Mixed commutators and little product BMO

Yumeng Ou, Stefanie Petermichl, Elizabeth Strouse

MIXED COMMUTATORS AND LITTLE PRODUCT BMO

YUMENG OU[1], STEFANIE PETERMICHL[2], AND ELIZABETH STROUSE

Abstract. We consider iterated commutators of multiplication by a symbol function and tensor products of Hilbert or Riesz transforms. We establish mixed BMO classes of symbols that characterize boundedness of these objects in L^p. Little BMO and product BMO, big Hankel operators and iterated commutators are the base cases of our results. We use operator theoretical methods and existing profound results on iterated commutators for the Hilbert transform case, while the general result in several variables is obtained through the construction of a Journé operator that models the behavior of the multiple Hilbert transform. Upper estimates for commutators with paraproduct free Journé operators as well as weak factorisation results are proven.

1. Introduction

A classical result of Nehari [24] shows that a Hankel operator with anti-analytic symbol b mapping analytic functions into the space of anti-analytic functions by $f \mapsto P_{-}bf$ is bounded with respect to an L^2 norm if and only if the symbol belongs to BMO. This theorem has an equivalent formulation in terms of the boundedness of the commutator of the multiplication operator with symbol function b and the Hilbert transform $[H, b] = Hb - bH$. To see this correspondence one rewrites the commutator as a sum of Hankel operators with orthogonal ranges.

Let $H^2(\mathbb{T}^2)$ denote the Banach space of analytic functions in $L^2(\mathbb{T}^2)$. In [13], Ferguson and Sadosky study the symbols of bounded 'big' and 'little' Hankel operators on the bidisk. Big Hankel operators are those which project on to a 'big' subspace of $L^2(\mathbb{T}^2)$ - the orthogonal complement of $H^2(\mathbb{T}^2)$; while little Hankel operators project onto the smaller subspace of complex conjugates of functions in $H^2(\mathbb{T}^2)$ - or anti-analytic functions. The corresponding commutators are

$$[H_1 H_2, b],$$

and

$$[H_1, [H_2, b]]$$

where $b = b(x_1, x_2)$ and H_k are the Hilbert transforms acting in the k^{th} variable. Ferguson and Sadosky show that the first commutator is bounded if and only if the symbol b belongs

2010 *Mathematics Subject Classification.* 42B20(Primary), 47B35(Secondary).

1. Research partly supported by NSF-DMS 0901139.
2. Research partly supported by ANR-12-BS01-0013-01. The author is a junior member of IUF.

to the so called little BMO class, consisting of those functions that are uniformly in BMO in each variable separately. They also show that if b belongs to the product BMO space, as identified by Chang and Fefferman [3], [4] then the second commutator is bounded. The fact that boundedness of the second commutator implies that b is in product BMO was shown in the groundbreaking paper of Ferguson and Lacey [12]. The techniques to tackle this question in several parameters are very different and have brought valuable new insight and use to existing theories, for example in the interpretation of Journé's lemma [16] in combination with Carleson's example [2]. Lacey and Terwilliger extended this result to an arbitrary number of iterates in [19], requiring thus, among others, a refinement of Pipher's iterated multi-parameter version [29] of Journé's lemma.

When leaving the notion of Hankel operators behind, their interpretation as commutators allow for natural generalizations. Through the use of completely different real variable methods, Coifman, Rochberg and Weiss [5] extended Nehari's one-parameter theory to real analysis in the sense that the Hilbert transforms were replaced by Riesz transforms. The missing features of the Riesz transforms include analytic projection on one hand as well as strong factorisation theorems of $H^1(\mathbb{D})$ on the other.

The authors in [5] obtained sufficiency, i.e. that a BMO symbol b yields an $L^2(\mathbb{R}^d)$ bounded commutator for certain more general, convolution type singular integral operators. For necessity, they showed that the collection of Riesz transforms was representative enough:

$$\|b\|_{\mathrm{BMO}} \lesssim \sup_{1\leqslant j\leqslant d} \|[R_j, b]\|_{2\to 2}.$$

Notably this lower bound was obtained somewhat indirectly through use of spherical harmonics in combination with the mean oscillation characterisation of BMO in one parameter.

These one-parameter results in [5] were extended to the multi-parameter setting in [17]. Both the upper and lower estimate have proofs very different from those in one parameter. For the lower estimate, the methods in [12] or [19] find an extension to real variables through operators closer to the Hilbert transform than the Riesz transforms and an indirect passage on the Fourier transform side.

In a recent paper [7] it is shown that iterated commutators formed with any arbitrary Calderón-Zygmund operators are bounded if the symbol belongs to product BMO.

The first part of the present note is concerned with mixed Hankel operators or commutators such as

$$[H_1, [H_2 H_3, b]].$$

We classify boundedness of these commutators by a mixed BMO class (little product BMO): those functions $b = b(x_1, x_2, x_3)$ so that $b(\cdot, x_2, \cdot)$ and $b(\cdot, \cdot, x_3)$ are uniformly in product BMO.

Similar results can be obtained for any finite iteration of any finite tensor product of Hilbert transforms.

The second part is concerned with a real variable analog or commutators of the form

$$[R_{1,j_1}, [R_{2,j_2} R_{3,j_3}, b]],$$

where R_{k,j_k} are Riesz transforms of direction j_k acting in the k^{th} variable. We show necessity and sufficiency of the little product BMO condition when the R_{k,j_k} are allowed to run through all Riesz transforms by means of a two-sided estimate. Our argument works for all higher iterates and tensor products.

After reading the formal definition of little product BMO, Definition 2.2, the main two-sided estimates are stated in Theorem 5.3.

It is a general fact that two-sided commutator estimates have an equivalent formulation in terms of weak factorization. We find the pre-duals of our little product BMO spaces and prove a corresponding weak factorization result.

Much like discussed in the base cases of our results [5], [17], boundedness of commutators involving Hilbert or Riesz transforms are a testing condition. If these commutators are bounded, the symbol necessarily belongs to a little product BMO. We then show that iterated commutators using a much more general class than that of tensor products of Riesz transforms are also bounded: commutators with paraproduct-free Journé operators.

We make some remarks about the strategy of the proof.

In the Hilbert transform case, Toeplitz operators with operator symbol arise naturally.

While Riesz transforms in $\mathbb{R}^d$ are a good generalisation of the Hilbert transform, there is absence of analytic structure and tools relying on analytic projection or orthogonal spaces are not readily available. We overcome this difficulty through a first intermediate passage via tensor products of Calderón-Zygmund operators whose Fourier multiplier symbols are adapted to cones. This idea is inspired by [17]. A class of operators of this type classifies little product BMO through two-sided commutator estimates, but it does not allow the passage to a classification through iterated commutators with tensor products of Riesz transforms. In a second step, we find it necessary to consider upper and lower commutator estimates using a well-chosen family of Journé operators that are not of tensor product type. These operators are constructed to resemble the multiple Hilbert transform. A two-sided estimate of iterated commutators involving operators of this family facilitates a passage to iterated commutators with tensor products of Riesz transforms. There is an increase in difficulty when the arising tensor products involve more than two Riesz transforms and when the dimension is greater than two.

The actual passage to the Riesz transforms requires for us to prove a stability estimate in commutator norms for certain multi-parameter singular integrals in terms of the mixed BMO class. In this context, we prove a qualitative upper estimate for iterated commutators using paraproduct free Journé operators. We make use of recent versions of $T(1)$ theorems in this setting. These recent advances are different from the corresponding theorem of Journé [15]. The results we allude to have the additional feature to provide a convenient representation formula for bi-parameter in [20] and even multi-parameter in [26] Calderón-Zygmund operators by dyadic shifts.

2. Aspects of Multi-Parameter Theory

This section contains some review on Hardy spaces in several parameters as well as some new definitions and lemmas relevant to us.

2.1. **Chang-Fefferman BMO.** We describe the elements of product Hardy space theory, as developed by Chang and Fefferman as well as Journé. By this we mean the Hardy spaces associated with domains like the poly-disk or $\mathbb{R}^{\vec{d}} := \bigotimes_{s=1}^{t} \mathbb{R}^{d_s}$ for $\vec{d} = (d_1, \ldots, d_t)$. While doing so, we typically do not distinguish whether we are working on $\mathbb{R}^d$ or $\mathbb{T}^d$. In higher dimensions, the Hilbert transform is usually replaced by the collections of Riesz transforms.

The (real) one-parameter Hardy space $H^1_{\mathrm{Re}}(\mathbb{R}^d)$ denotes the class of functions with the norm

$$\sum_{j=0}^{d} \|R_j f\|_1$$

where R_j denotes the j^{th} Riesz transform or the Hilbert transform if the dimension is one. Here and below we adopt the convention that R_0, the 0^{th} Riesz transform, is the identity. This space is invariant under the one-parameter family of isotropic dilations, while the product Hardy space $H^1_{\mathrm{Re}}(\mathbb{R}^{\vec{d}})$ is invariant under dilations of each coordinate separately. That is, it is invariant under a t parameter family of dilations, hence the terminology 'multi-parameter' theory. One way to define a norm on $H^1_{\mathrm{Re}}(\mathbb{R}^{\vec{d}})$ is

$$\|f\|_{H^1} \sim \sum_{0 \leqslant j_l \leqslant d_l} \|\bigotimes_{l=1}^{t} \mathrm{R}_{l,j_l} f\|_1.$$

R_{l,j_l} is the Riesz transform in the j_l^{th} direction of the l^{th} variable, and the 0^{th} Riesz transform is the identity operator.

The dual of the real Hardy space $H^1_{\mathrm{Re}}(\mathbb{R}^{\vec{d}})^*$ is $\mathrm{BMO}(\mathbb{R}^{\vec{d}})$, the t-fold product BMO space. It is a theorem of S.-Y. Chang and R. Fefferman [3], [4] that this space has a characterization in terms of a product Carleson measure.

Define

$$\|b\|_{\mathrm{BMO}(\mathbb{R}^{\vec{d}})} := \sup_{U\subset\mathbb{R}^{\vec{d}}} \left(|U|^{-1} \sum_{R\subset U} \sum_{\vec{\varepsilon}\in\mathrm{sig}_{\vec{d}}} |\langle b, w_R^{\vec{\varepsilon}}\rangle|^2 \right)^{1/2}. \tag{1}$$

Here the supremum is taken over all open subsets $U \subset \mathbb{R}^{\vec{d}}$ with finite measure, and we use a wavelet basis $w_R^{\vec{\varepsilon}}$ adapted to rectangles $R = Q_1 \times \cdots \times Q_t$, where each Q_l is a cube. The superscript $\vec{\varepsilon}$ reflects the fact that multiple wavelets are associated to any dyadic cube, see [17] for details. The fact that the supremum admits all open sets of finite measure cannot be omitted, as Carleson's example shows [2]. This fact is responsible for some of the difficulties encountered when working with this space.

Theorem 2.1 (Chang, Fefferman)**.** *We have the equivalence of norms*

$$\|b\|_{(H^1_{Re}(\mathbb{R}^{\vec{d}}))^*} \sim \|b\|_{BMO(\mathbb{R}^{\vec{d}})}.$$

That is, $BMO(\mathbb{R}^{\vec{d}})$ is the dual to $H^1_{Re}(\mathbb{R}^{\vec{d}})$.

This BMO norm is invariant under a t-parameter family of dilations. Here the dilations are isotropic in each parameter separately. See also [9] and [11].

2.2. **Little BMO.** Following [6] and [13], we review the space little BMO, often written as 'bmo', and its predual. A locally integrable function $b : \mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_s} \to \mathbb{C}$ is in bmo if and only if

$$\|b\|_{\mathrm{bmo}} = \sup_{\vec{Q}=Q_1\times\cdots\times Q_s} |\vec{Q}|^{-1} \int_{\vec{Q}} |b(\vec{x}) - b_{\vec{Q}}| < \infty$$

Here the Q_k are d_k-dimensional cubes and $b_{\vec{Q}}$ denotes the average of b over $\vec{Q}$.

It is easy to see that this space consists of all functions that are uniformly in BMO in each variable separately. Let $\vec{x}_{\hat{v}} = (x_1, \ldots, x_{v-1}, \cdot, x_{v+1}, \ldots, x_s)$. Then $b(\vec{x}_{\hat{v}})$ is a function in x_v only with the other variables fixed. Its BMO norm in x_v is

$$\|b(\vec{x}_{\hat{v}})\|_{\mathrm{BMO}} = \sup_{Q_v} |Q_v|^{-1} \int_{Q_v} |b(\vec{x}) - b(\vec{x}_{\hat{v}})_{Q_v}| dx_v$$

and the little BMO norm becomes

$$\|b\|_{\mathrm{bmo}} = \max_v \{\sup_{\vec{x}_{\hat{v}}} \|b(\vec{x}_{\hat{v}})\|_{\mathrm{BMO}}\}.$$

On the bi-disk, this becomes

$$\|b\|_{\mathrm{bmo}} = \max\{\sup_{x_1} \|b(x_1, \cdot)\|_{\mathrm{BMO}}, \sup_{x_2} \|b(\cdot, x_2)\|_{\mathrm{BMO}}\},$$

the space discussed in [13]. Here, the pre-dual is the space $H^1(\mathbb{T}) \otimes L^1(\mathbb{T}) + L^1(\mathbb{T}) \otimes H^1(\mathbb{T})$. All other cases are an obvious generalisation, at the cost of notational inconvenience.

2.3. **Little product BMO.** In this section we define a BMO space which is in between little BMO and product BMO. As mentioned in the introduction, we aim at characterising BMO spaces consisting for example of those functions $b(x_1, x_2, x_3)$ such that $b(x_1, \cdot, \cdot)$ and $b(\cdot, \cdot, x_3)$ are uniformly in product BMO in the remaining two variables.

Definition 2.2. *Let $b : \mathbb{R}^{\vec{d}} \to \mathbb{C}$ with $\vec{d} = (d_1, \cdots, d_t)$. Take a partition $\mathcal{I} = \{I_s : 1 \leqslant s \leqslant l\}$ of $\{1, 2, ..., t\}$ so that $\dot\cup_{1 \leqslant s \leqslant l} I_s = \{1, 2, ..., t\}$. We say that $b \in BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ if for any choices $\boldsymbol{v} = (v_s), v_s \in I_s$, b is uniformly in product BMO in the variables indexed by v_s. We call a BMO space of this type a 'little product BMO'. If for any $\vec{x} = (x_1, ..., x_t) \in \mathbb{R}^{\vec{d}}$, we define $\vec{x}_{\hat{\boldsymbol{v}}}$ by removing those variables indexed by v_s, the little product BMO norm becomes*

$$\|b\|_{BMO_{\mathcal{I}}} = \max_{\boldsymbol{v}} \{\sup_{\vec{x}_{\hat{\boldsymbol{v}}}} \|b(\vec{x}_{\hat{\boldsymbol{v}}})\|_{BMO}\}$$

where the BMO norm is product BMO in the variables indexed by v_s.

For example, when $\vec{d} = (1, 1, 1) = \vec{1}$, when $t = 3$ and $l = 2$ with $I_1 = (13)$ and $I_2 = (2)$, writing $\mathcal{I} = (13)(2)$ the space $\mathrm{BMO}_{(13)(2)}(\mathbb{T}^{\vec{1}})$ arises, which consists of those functions that are uniformly in product BMO in the variables $(1, 2)$ and $(3, 2)$ respectively, as described above. Some other examples of little product BMOs are $\mathrm{BMO}_{(12)(34)}$ and $\mathrm{BMO}_{(234)(1)(5)}$. The first space consists of those functions which are uniformly in product BMO in the variables $(1, 3)$, $(1, 4)$, $(2, 3)$ and $(2, 4)$, and the second one consists of functions that are uniformly in product BMO in the variables $(2, 1, 5)$, $(3, 1, 5)$ and $(4, 1, 5)$. Moreover, as degenerate cases, it is easy to see that $\mathrm{BMO}_{(12...t)}$ and $\mathrm{BMO}_{(1)(2)...(t)}$ are exactly little BMO and product BMO respectively, the spaces we are familiar with.

The little product BMO spaces are closely related to classes of iterated commutators of tensors of operators, a fact we will elaborate on throughout the paper.

Little product BMO spaces on $\mathbb{T}^{\vec{d}}$ can be defined in the same way. Now we find the predual of $\mathrm{BMO}_{(13)(2)}$, which is a good model for other cases. We choose the order of variables most convenient for us.

Theorem 2.3. *The pre-dual of the space $BMO_{(13)(2)}(\mathbb{T}^{\vec{1}})$ is equal to the space*

$$\begin{aligned}&H^1_{Re}(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T}) + L^1(\mathbb{T}) \otimes H^1_{Re}(\mathbb{T}^{(1,1)}) \\ &:= \{f + g : f \in H^1_{Re}(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T}) \text{ and } g \in L^1(\mathbb{T}) \otimes H^1_{Re}(\mathbb{T}^{(1,1)})\}.\end{aligned}$$

Proof. The space

$$H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T}) = \{f \in L^1(\mathbb{T}^3) : H_1 f, H_2 f, H_1 H_2 f \in L^1(\mathbb{T}^3)\}$$

equipped with the norm $\|f\| = \|f\|_1 + \|H_1 f\|_1 + \|H_2 f\|_1 + \|H_1 H_2 f\|_1$ is a Banach space. Let $W^1 = L^1(\mathbb{T}^3) \times L^1(\mathbb{T}^3) \times L^1(\mathbb{T}^3) \times L^1(\mathbb{T}^3)$ equipped with the norm

$$\|(f_1, f_2, f_3, f_4)\|_{W_1} = \|f_1\|_1 + \|f_2\|_1 + \|f_3\|_1 + \|f_4\|_1.$$

Then we see that $H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T})$ is isomorphically isometric to the closed subspace

$$V = \{(f, H_1(f), H_2(f), H_1H_2(f)) : f \in H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T})\}$$

of W^1. Now, the dual of W^1 is equal to $W^\infty = L^\infty(\mathbb{T}^3) \times L^\infty(\mathbb{T}^3) \times L^\infty(\mathbb{T}^3) \times L^\infty(\mathbb{T}^3)$ equipped with the norm $\|(g_1, g_2, g_3, g_4)\|_\infty = \max\{\|g_i\|_\infty (i = 1, \cdots 4)\}$ so the dual space of V is equal to the quotient of W^∞ by the annihilator U of the subspace V in W^∞. But, using the fact that the Hilbert transforms are self-adjoint up to a sign change, we see that

$$U = \{(g_1, g_2, g_3, g_4) : g_1 + H_1g_2 + H_2g_3 + H_1H_2g_4 = 0\}$$

and so:

$$V^* \cong W^\infty/U \cong \mathrm{Im}(\theta)$$

where

$$\theta(g_1, g_2, g_3, g_4) = g_1 + H_1g_2 + H_2g_3 + H_1H_2g_4$$

since $U = \ker(\theta)$. But

$$\mathrm{Im}(\theta) = L^\infty(\mathbb{T}^3) + H_1(L^\infty(\mathbb{T}^3)) + H_2(L^\infty(\mathbb{T}^3)) + H_1(H_2(L^\infty(\mathbb{T}^3)))$$

is equal to the functions that are uniformly in product BMO in variables 1 and 2.

Using the same reasoning we see that the dual of $L^1(\mathbb{T}) \otimes H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)})$ is equal to $L^\infty(\mathbb{T}^3) + H_2(L^\infty(\mathbb{T}^3)) + H_3(L^\infty(\mathbb{T}^3)) + H_2H_3(L^\infty(\mathbb{T}^3))$, which is equal to the space of functions that are uniformly in product BMO in variable 2 and 3.

Now, we consider the 'L^1 sum' of the spaces $H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T})$ and $L^1(\mathbb{T}) \otimes H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)})$; that is

$$M_{(13)(2)} = \{(f, g) : f \in H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T}); g \in L^1(\mathbb{T}) \otimes H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)})\}$$

equipped with the norm

$$\|(f, g)\| = \|f\|_{H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)})\otimes L^1(\mathbb{T})} + \|g\|_{L^1(\mathbb{T})\otimes H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)})}.$$

We see that, if $\phi : M_{(13)(2)} \to L^1((\mathbb{T}^3)$ is defined by $\phi(f, g) = f + g$, then the image of ϕ is isometrically isomorphic to the quotient of $M_{(13)(2)}$ by the space

$$N = \{(f, g) \in M_{(13)(2)} : f+g = 0\} = \{(f, -f) : f \in H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)})\otimes L^1(\mathbb{T}) \cap L^1(\mathbb{T})\otimes H^1_{\mathrm{Re}}(\mathbb{T}^{(1,1)})\}.$$

Now, recall that the dual of the quotient M/N is equal to the annihilator of N. It is easy to see that the annihilator of N is equal to the set of ordered pairs (ϕ, ϕ) with ϕ in the intersection of the duals of the two spaces. Thus the dual of the image of θ is equal to $\mathrm{BMO}_{(13)(2)}$. The norm of an element in the predual is equal to its norm as an element of the double dual which is easily computed. □

Following this example, the reader may easily find the correct formulation for the predual of other little product BMO spaces as well those in several variables, replacing the Hilbert transform by all choices of Riesz transforms. For instance, one can prove that the predual of the space $\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})$ is equal to $H^1_{\mathrm{Re}}(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1_{\mathrm{Re}}(\mathbb{R}^{(d_2,d_3)})$.

3. The Hilbert transform case

In this section, we characterise the boundedness of commutators of the form $[H_2, [H_3H_1, b]]$ as operators on $L^2(\mathbb{T}^3)$. In the case of the Hilbert transform, this case is representative for the general case and provides a starting point that is easier to read because of the simplicity of the expression of products and sums of projection onto orthogonal subspaces. Its general form can be found at the beginning of Section 5.

Now let $b \in L^1(\mathbb{T}^n)$ and let P and Q denote orthogonal projections onto subspaces of $L^2(\mathbb{T}^n)$. We shall describe relationships between functions in the little product BMOs and several types of projection-multiplication operators. These will be Hilbert transform-type operators of the form $P - P^\perp$; and iterated Hankel or Toeplitz type operators of the form $Q^\perp bQ$ (Hankel), PbP (Toeplitz),$PQ^\perp bQP$ (mixed), where b means the (not a priori bounded) multiplication operator M_b on $L^2(\mathbb{T}^n)$.

We shall use the following simple observation concerning Hilbert transform type operators again and again:

Remark 3.1. *If $H = P - P^\perp$ and $T : L^2(\mathbb{T}^n) \to L^2(\mathbb{T}^n)$ is a linear operator then*

$$[H, T] = 2PTP^\perp - 2P^\perp TP$$

and H is bounded if and only if $PTP^\perp$ and $P^\perp TP$ are.

Proof.

$$\begin{aligned}(P - P^\perp)T - T(P - P^\perp) &= (P - P^\perp)T(P + P^\perp) - (P + P^\perp)T(P - P^\perp)\\ &= 2PTP^\perp - 2P^\perp TP.\end{aligned}$$

The observation on boundedness comes from the fact that H is expressed as the sum of operators with orthogonal ranges. □

We state the main result of this section.

Theorem 3.2. *Let $b \in L^1(\mathbb{T}^3)$. Then the following are equivalent with linear dependence on the respective norms*

(1) $b \in BMO_{(13)(2)}$
(2) The commutators $[H_2, [H_1, b]]$ and $[H_2, [H_3, b]]$ are bounded on $L^2(\mathbb{T}^3)$
(3) The commutator $[H_2, [H_3H_1, b]]$ is bounded on $L^2(\mathbb{T}^3)$.

Corollary 3.3. *We have the following two-sided estimate*

$$\|b\|_{BMO_{(13)(2)}} \lesssim \|[H_2, [H_3H_1, b]]\|_{L^2(\mathbb{T}^3)\to L^2(\mathbb{T}^3)} \lesssim \|b\|_{BMO_{(13)(2)}}.$$

Before we proceed with the proof of Theorem 3.2, we are going to need some notation and some lemmas. It will be useful to denote by Q_{13} orthogonal projection on the subspace of functions which are either analytic or anti-analytic in the first and third variables; $Q_{13} = P_1P_3 + P_1^{\perp}P_3^{\perp}$. Then the projection $Q_{13}^{\perp}$ onto the orthogonal of this subspace is defined by $Q_{13}^{\perp} = P_1^{\perp}P_3 + P_1P_3^{\perp}$. We reformulate properties *(2)* and *(3)* in the statement of Theorem 3.2 in terms of Hankel Toeplitz type operators.

Lemma 3.4. *We have the following algebraic facts on commutators and projection operators.*

(1) The commutators $[H_2,[H_1,b]]$ and $[H_2,[H_3,b]]$ are bounded on $L^2(\mathbb{T}^3)$ if and only if $P_iP_2bP_i^{\perp}P_2^{\perp}$, $P_i^{\perp}P_2bP_iP_2^{\perp}$, $P_iP_2^{\perp}bP_i^{\perp}P_2$, $P_i^{\perp}P_2^{\perp}bP_iP_2$ with $i \in \{1,3\}$ are bounded on $L^2(\mathbb{T}^3)$.

(2) The commutator $[H_2,[H_3H_1,b]]$ is bounded on $L^2(\mathbb{T}^3)$ if and only if all four iterated Hankels $P_2Q_{13}bQ_{13}^{\perp}P_2^{\perp}$, $P_2^{\perp}Q_{13}^{\perp}bQ_{13}P_2$, $P_2Q_{13}^{\perp}bQ_{13}P_2^{\perp}$, $P_2^{\perp}Q_{13}bQ_{13}^{\perp}P_2$ are bounded on $L^2(\mathbb{T}^3)$.

Proof. Using Remark 3.1 it is easy to see that

$$[H_2,[H_1,b]] = 4\big((P_2P_1bP_1^{\perp}P_2^{\perp} - P_2P_1^{\perp}bP_1P_2^{\perp}) - (P_2^{\perp}P_1bP_1^{\perp}P_2 - P_2^{\perp}P_1^{\perp}bP_1P_2)\big)$$

and that the corresponding equation for $[H_2,[H_3,b]]$ is also true. This, along with the observation that the ranges of all arising summands are mutually orthogonal, so that these operators are bounded if and only if each one of their summands is, gives assertion *(1)*. To prove *(2)* we just notice that $H_1H_3 = Q_{13} - Q_{13}^{\perp}$ is one of our Hilbert transform type operators which permits us to calculate:

$$[H_2,[H_3H_1,b]] = 4\big((P_2Q_{13}bQ_{13}^{\perp}P_2^{\perp} - P_2Q_{13}^{\perp}bQ_{13}P_2^{\perp}) - (P_2^{\perp}Q_{13}bQ_{13}^{\perp}P_2 - P_2^{\perp}Q_{13}bQ_{13}^{\perp}P_2)\big)$$

and conclude that the commutator is bounded if and only if the iterated Hankels are. □

The other tools that we need in order to prove *(3)* ⇒ *(2)*. in Theorem 3.2 are Lemmas 3.5 and 3.6, Toeplitz type Lemmas which use the fact that norms of Toeplitz operators are determined by the L^{∞} norms of their symbols.

Lemma 3.5. *Let W_3 be the operator of multiplication by z_3, $W_3(f) = z_3f$, acting on $L^2(\mathbb{T}^3)$. If we define $B = P_1^{\perp}P_2^{\perp}bP_1P_2$ as well as*

$$A_n = W_3^{*n}(P_3P_1^{\perp}P_2^{\perp}bP_1P_2P_3)W_3^n \text{ and } C_n = W_3^n(P_3^{\perp}P_1^{\perp}P_2^{\perp}bP_1P_2P_3^{\perp})W_3^{*n}$$

as operators acting on $L^2(\mathbb{T}^3)$ then the sequences A_n and C_n converge to B in the strong operator topology.

The proof below is a generalization of the Brown-Halmos classical proof that, if S is the unilateral shift and T_b is the Toeplitz operator with symbol b on $L^2(\mathbb{T})$, then $S^{*n}T_bS^n$ converges in the strong operator topology to M_b the multiplication operator with symbol b [1]. We sketch the argument adapted to our situation.

Proof. The two assertions are symmetric, we prove the fact for A_n. It is easy to see that W_3 , W_3^*; and P_3 commute with $P_1, P_2, P_1^\perp$ and $P_2^\perp$. The multiplier b satisfies the equation $W_3^{*n} b W_3^n = b$ and $W_3^n W_3^{*n} = Id$. So we see that

$$A_n = P_1^\perp P_2^\perp (W_3^{*n} P_3 W_3^n) b P_1 P_2 (W_3^{*n} P_3 W_3^n).$$

But if $f \in L^2(\mathbb{T}^3)$, then, since W_3^n is a unitary operator:

$$\|W_3^{*n} P_3 W_3^n(f) - f\| = \|P_3 W_3^n(f) - W_3^n(f)\| = \|(P_3 - I)(W_3^n)(f)\| \to 0 \quad (n \to \infty),$$

as tail of a convergent Fourier series. This means that

$$W_3^{*n} P_3 W_3^n : L^2(\mathbb{T}^3) \to L^2(\mathbb{T}^3)$$

converges to the identity in the strong operator topology. Thus, for each $f \in L^2(\mathbb{T}^3)$ we have

$$\begin{aligned} \|(A_n - B)(f)\| &\leqslant \|(P_1^\perp P_2^\perp (W_3^{*n} P_3 W_3^n - Id) b P_1 P_2 W_3^{*n} P_3 W_3^n)(f)\| \\ &\quad + \|(P_1^\perp P_2^\perp b P_1 P_2 (W_3^{*n} P_3 W_3^n - Id))(f)\| \to 0 \quad (n \to \infty) \end{aligned}$$

□

Lemma 3.6. $\|P_3 P_1^\perp P_2^\perp b P_1 P_2 P_3\|_{L^2 \to L^2} = \|P_1^\perp P_2^\perp b P_1 P_2\|_{L^2 \to L^2}$.

Proof. The inequality $\leqslant$ is trivial, since P_3 is a projection which commutes with $P_1^\perp$ and $P_2^\perp$.

By Lemma 3.5,

$$\begin{aligned} \|P_1^\perp P_2^\perp b P_1 P_2\| &\leqslant \sup_{n \in \mathbb{N}} \|W_3^{*n} (P_3 P_1^\perp P_2^\perp b P_1 P_2 P_3) W_3^n\| \\ &\leqslant \|P_3 P_1^\perp P_2^\perp b P_1 P_2 P_3\|, \end{aligned}$$

using again the fact that the operators W_3 and W_3^* are isometries on $L^2(\mathbb{T}^3)$. □

Now, we are ready to proceed with the proof of the main theorem of this section.

Proof. (of Theorem 3.2) We show *(1)* $\Leftrightarrow$ *(2)* and *(2)* $\Leftrightarrow$ *(3)*

(1) $\Leftrightarrow$ *(2)* Consider $f = f(x_1, x_2)$ and $g = g(x_3)$. Observe that $[H_2, [H_1, b]](fg) = g \cdot [H_2, [H_1, b]](f)$. So $\|[H_2, [H_1, b]](fg)\|^2_{L^2(\mathbb{T}^3)} = \|Fg\|^2_{L^2(\mathbb{T})}$ where $F(x_2) = \|[H_2, [H_1, b]](f)\|_{L^2(\mathbb{T}^2)}$. The map $g \mapsto Fg$ has $L^2(\mathbb{T})$ operator norm $\|F\|_\infty$. Now change the roles of x_1 and x_3. The Ferguson-Lacey equivalences $\|[H_2, [H_i, b]]\| \sim \|b\|_{\mathrm{BMO}}$ give the desired result.

(2) $\Rightarrow$ *(3)* We show that boundedness of the commutators $[H_2, [H_1, b]]$ and $[H_2, [H_3, b]]$ implies the boundedness of the mixed commutator $[H_2, [H_1 H_3, b]]$. Indeed, the boundedness of the mixed commutator is equivalent to the boundedness of the four operators listed in Lemma 3.4. By direct calculation we can transform, for example, the first term $P_2 Q_{13} b Q_{13}^\perp P_2^\perp$

into the sum of four terms with orthogonal range:

$$\begin{aligned} P_2Q_{13}bQ_{13}^{\perp}P_2^{\perp} =& P_3(P_1P_2bP_1^{\perp}P_2^{\perp})P_3 + P_3^{\perp}(P_1^{\perp}P_2bP_1P_2^{\perp})P_3^{\perp} \\ &+ P_1(P_2P_3bP_2^{\perp}P_3^{\perp})P_1 + P_1^{\perp}(P_2P_3^{\perp}bP_2^{\perp}P_3)P_1^{\perp}. \end{aligned}$$

Notice that each term of this sum is either of the form P_iTP_i or of the form $P_i^{\perp}TP_i^{\perp}$ where T is one of the eight operators from Lemma 3.4 which are bounded because we assume *(2)*. Clearly $\|P_iTP_i\| \leqslant \|T\|$ and $\|P_i^{\perp}TP_i^{\perp}\| \leqslant \|T\|$ and so $P_2Q_{13}bQ_{13}^{\perp}P_2^{\perp}$ is bounded. Analogous equations hold for the other four operators of *(3)* and so $[H_2,[H_3H_1,b]]$ is bounded.

(3) $\Rightarrow$ *(2)* This part relies on Lemma 3.6, which in turn relies on Lemma 3.5. We wish to conclude from the boundedness of $[H_2,[H_3H_1,b]]$ the boundedness of $[H_2,[H_1,b]]$ and $[H_2,[H_3,b]]$. To see boundedness of $[H_2,[H_1,b]]$, let us look at one of the Hankels from Lemma 3.4. Lemma 3.6 shows that $P_2^{\perp}P_1^{\perp}bP_2P_1$ is bounded if and only if the operator $P_3P_1^{\perp}P_2^{\perp}bP_1P_2P_3$ is. And the latter is an operator found in the list from part *(2)* of Lemma 3.4. The analogous reasoning shows that all eight Hankels in 3.4 are bounded and so *2* is proved. □

4. Real variables: upper bounds

In this section, we are interested in upper bounds for commutator norms by means of little product BMO norms of the symbol. In the case of the Hilbert transform, we have seen that these estimates, even in the iterated case, are straightforward. Other streamlined proofs exist for Hilbert or Riesz transforms when considering dyadic shifts of complexity one, see [27], [28] and [18]. When considering more general Calderón-Zygmund operators, the arguments required are more difficult, in each case. The first classical upper bound goes back to [5], where an estimate for one-parameter commutators with convolution type Calderón-Zygmund operators is given. Next, the text [17] includes a technical estimate for the multi-parameter case for such Calderón-Zygmund operators with a high enough degree of smoothness. This smoothness assumption was removed in [7] thanks to an approach using the representation formula for Calderón-Zygmund operators by means of infinite complexity dyadic shifts [14]. This last proof also gives a control on the norm of the commutators which dependends on the Calderón-Zygmund norm of the operators themselves, a fact we will employ later. Below, we prove two generalizations to the mixed commutator setting. The first theorem is weaker and targets the tensor product case only. It will be relevant to the L^p case in Section 7. The second theorem gives an estimate by little product BMO when the Calderón-Zygmund operators are of Journé type and cannot be written as a tensor product. While this estimate is interesting in its own right, it is also essential for our characterisation result, the lower estimate, in section 5. We start with the tensor product case, which is an almost direct consequence of [7].

Theorem 4.1. *Let $\vec{d} = (d_1, \ldots, d_t)$ and $\mathcal{I} = (I_s)_{1 \leqslant s \leqslant l}$ a partition of $\{1, \ldots, t\}$. For each $1 \leqslant s \leqslant l$ let us consider the tensor product of Calderón-Zygmund operators $\vec{T}_s = \bigotimes_{k \in I_s} T_k$ where T_k acts on functions defined on the k^{th} variable. Let the symbol b belong to the little product BMO space $BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$. Then,*

$$\|[\vec{T}_1, \ldots [\vec{T}_l, b] \ldots]\|_{L^2(\mathbb{R}^{\vec{d}}) \circlearrowleft} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})},$$

where the underlying constant depends only on dimensions and $\|T_k\|_{CZ}$.

The main ingredient of the proof is the upper bound theorem in [7]. Moreover, the one-parameter result (without iteration) can be similarly obtained, which we omit.

Proof. For convenience of notation, we provide the proof in the case $\mathbb{R}^{(d_1, d_2, d_3)}$ with partition $\mathcal{I}$ consisting of $I_1 = (1)$ and $I_2 = (23)$ and the commutator $[\vec{T}_1, [\vec{T}_2, b]] = [T_1, [T_2 T_3, b]]$ only. The argument is simple and generalizes easily. By the fact that $[T_2 T_3, b] = T_2[T_3, b] + [T_2, b]T_3$, one has

$$\begin{aligned}
[T_1, [T_2 T_3, b]] &= [T_1, T_2[T_3, b]] + [T_1, [T_2, b]T_3] \\
&= T_2[T_1, [T_3, b]] + [T_1, [T_2, b]]T_3.
\end{aligned}$$

We consider the first term only, the second term can be dealt with in the same way.

$$\begin{aligned}
\|T_2[T_1, [T_3, b]]f\|^2_{L^2(\mathbb{R}^{\vec{d}})} &\lesssim \|[T_1, [T_3, b]]f\|^2_{L^2(\mathbb{R}^{\vec{d}})} \\
&= \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_3}} |[T_1, [T_3, b]]f|^2 \, dx_1 dx_3 \right) dx_2 \\
&\leqslant \int_{\mathbb{R}^{d_2}} \|[T_1, [T_3, b(\cdot, x_2, \cdot)]]\|^2_{L^2(\mathbb{R}^{d_1} \times \mathbb{R}^{d_3}) \circlearrowleft} \|f(\cdot, x_2 \cdot)\|^2_{L^2(\mathbb{R}^{d_1} \times \mathbb{R}^{d_3})} \, dx_2 \\
&\lesssim \int_{\mathbb{R}^{d_2}} \|b(\cdot, x_2, \cdot)\|^2_{\mathrm{BMO}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_3})} \|f(\cdot, x_2, \cdot)\|^2_{L^2(\mathbb{R}^{d_1} \times \mathbb{R}^{d_3})} \, dx_2 \\
&\leqslant \|b\|^2_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \|f\|^2_{L^2(\mathbb{R}^{\vec{d}})}.
\end{aligned}$$

The implied constants depend upon the L^2 norm of T_2 in the first inequality and the Calderón-Zygmund norms of T_1 and T_3 in the fourth line. Indeed, for any fixed x_2, by the main theorem of [7], $\|[T_1, [T_3, b(\cdot, x_2, \cdot)]]\|_{L^2(\mathbb{R}^{d_1} \times \mathbb{R}^{d_3}) \circlearrowleft} \lesssim \|b(\cdot, x_2, \cdot)\|_{\mathrm{BMO}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_3})}$ with the implied constant depending only on the dimensions and $\|T_1\|_{CZ}, \|T_3\|_{CZ}$. The last inequality follows from $\|b(\cdot, x_2, \cdot)\|_{\mathrm{BMO}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_3})} \leqslant \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}$ for all x_2. □

The first generation of multi-parameter singular integrals that are not of tensor product type goes back to Fefferman [10] and was generalised by Journé in [15] to the non-convolution type in the framework of his $T(1)$ theorem in this setting. Much later, Journé's $T(1)$ theorem was revisited, for example in [30], [20], [25], [26]. See also [21] for some difficulties related to this subject. The references [20] in the bi-parameter case and [26] in the general multi-parameter case include a representation formula by means of adapted, infinite complexity dyadic shifts. While these representation formulae look complicated, they have a feature very useful to us. Locally' they look as if they were of tensor product type, a feature we will exploit in the argument below. We start with the simplest bi-parameter case with no iterations and make comments about the generalization at the end of this section.

The class of bi-parameter singular integral operators treated in this section is that of any paraproduct free Journé type operator (not necessarily a tensor product and not necessarily of convolution type) satisfying a certain weak boundedness property, which we define as follows:

Definition 4.2. *A continuous linear mapping* $T : C_0^\infty(\mathbb{R}^n)\otimes C_0^\infty(\mathbb{R}^m) \to [C_0^\infty(\mathbb{R}^n)\otimes C_0^\infty(\mathbb{R}^m)]'$ *is called a* paraproduct free bi-parameter Calderón-Zygmund operator *if the following conditions are satisfied:*

1. T is a Journé type bi-parameter δ-singular integral operator, i.e. there exists a pair (K_1, K_2) of δCZ-δ-standard kernels so that, for all $f_1, g_1 \in C_0^\infty(\mathbb{R}^n)$ and $f_2, g_2 \in C_0^\infty(\mathbb{R}^m)$,

$$\langle T(f_1 \otimes f_2), g_1 \otimes g_2\rangle = \int f_1(y_1)\langle K_1(x_1, y_1)f_2, g_2\rangle g_1(x_1)\, dx_1dy_1$$

when $sptf_1 \cap sptg_1 = \varnothing$;

$$\langle T(f_1 \otimes f_2), g_1 \otimes g_2\rangle = \int f_2(y_2)\langle K_2(x_2, y_2)f_1, g_1\rangle g_2(x_2)\, dx_2dy_2$$

when $sptf_2 \cap sptg_2 = \varnothing$.

2. T satisfies the weak boundedness property $|\langle T(\chi_I \otimes \chi_J), \chi_I \otimes \chi_J\rangle| \lesssim |I||J|$, for any cubes $I \subset \mathbb{R}^n$, $J \in \mathbb{R}^m$.

3. T is paraproduct free in the sense that $T(1\otimes \cdot) = T(\cdot \otimes 1) = T^*(1\otimes \cdot) = T^*(\cdot \otimes 1) = 0$.

Recall that δCZ-δ-standard kernel is a vector valued standard kernel taking values in the Banach space consisting of all Calderón-Zygmund operators. It is easy to see that an operator defined as above satisfies all the characterizing conditions in Martikainen [20], hence is L^2 bounded and can be represented as an average of bi-parameter dyadic shift operators together with dyadic paraproducts. Moreover, since T is paraproduct free, one can conclude from observing the proof of Martikainen's theorem, that all the dyadic shifts in the representation are cancellative.

The base case from which we pass to the general case below, is the following:

Theorem 4.3. *Let T be a paraproduct free bi-parameter Calderón-Zygmund operator, and b be a little bmo function, there holds*

$$\|[b, T]\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)\circlearrowleft} \lesssim \|b\|_{bmo(\mathbb{R}^n\times\mathbb{R}^m)},$$

where the underlying constant depends only on the characterizing constants of T.

Proof. According to the discussion above, for any sufficiently nice functions f, g, one has the following representation:

$$\langle Tf, g\rangle = C\mathbb{E}_{\omega_1}\mathbb{E}_{\omega_2} \sum_{i_1,j_1=0}^{\infty} \sum_{i_2,j_2=0}^{\infty} 2^{-\max(i_1,j_1)} 2^{-\max(i_2,j_2)} \langle S^{i_1j_1i_2j_2} f, g\rangle. \tag{2}$$

where expectation is with respect to a certain parameter of the dyadic grids. The dyadic shifts $S^{i_1j_1i_2j_2}$ are defined as

$$\begin{aligned} S^{i_1j_1i_2j_2} f &:= \sum_{K_1\in\mathcal{D}_1} \sum_{\substack{I_1,J_1\subset K_1, I_1,J_1\in\mathcal{D}_1\\ \ell(I_1)=2^{-i_1}\ell(K_1)\\ \ell(J_1)=2^{-j_1}\ell(K_1)}} \sum_{K_2\in\mathcal{D}_2} \sum_{\substack{I_2,J_2\subset K_2, I_2,J_2\in\mathcal{D}_2\\ \ell(I_2)=2^{-i_2}\ell(K_2)\\ \ell(J_2)=2^{-j_2}\ell(K_2)}} a_{I_1J_1K_1I_2J_2K_2}\langle f, h_{I_1}\otimes h_{I_2}\rangle h_{J_1}\otimes h_{J_2} \\ &=: \sum_{K_1} \sum_{I_1,J_1\subset K_1}^{(i_1,j_1)} \sum_{K_2} \sum_{I_2,J_2\subset K_2}^{(i_2,j_2)} a_{I_1J_1K_1I_2J_2K_2}\langle f, h_{I_1}\otimes h_{I_2}\rangle h_{J_1}\otimes h_{J_2}. \end{aligned}$$

The coefficients above satisfy $a_{I_1J_1K_1I_2J_2K_2} \leqslant \frac{\sqrt{|I_1||J_1||I_2||J_2|}}{|K_1||K_2|}$, which also guarantees that $\|S^{i_1j_1i_2j_2}\|_{L^2\to L^2} \leqslant 1$. Moreover, since T is paraproduct free, all the Haar functions appearing above are cancellative.

It thus suffices to show that for any dyadic grids $\mathcal{D}_1, \mathcal{D}_2$ and fixed $i_1, j_1, i_2, j_2 \in \mathbb{N}$, one has

$$\|[b, S^{i_1j_1i_2j_2}]f\|_{L^2} \lesssim (1+\max(i_1,j_1))(1+\max(i_2,j_2))\|b\|_{\text{bmo}}\|f\|_{L^2}, \tag{3}$$

as the decay factor $2^{-\max(i_1,j_1)}, 2^{-\max(i_2,j_2)}$ in (2) will guarantee the convergence of the series.

To see (3), one decomposes b and a L^2 test function f using Haar bases:

$$[b, S^{i_1j_1i_2j_2}]f = \sum_{I_1,I_2}\sum_{J_1,J_2} \langle b, h_{I_1}\otimes h_{I_2}\rangle\langle f, h_{J_1}\otimes h_{J_2}\rangle [h_{I_1}\otimes h_{I_2}, S^{i_1j_1i_2j_2}] h_{J_1}\otimes h_{J_2}.$$

A similar argument to that in [7] implies that $[h_{I_1}\otimes h_{I_2}, S^{i_1j_1i_2j_2}]h_{J_1}\otimes h_{J_2}$ is nonzero only if $I_1 \subset J_1^{(i_1)}$ or $I_2 \subset J_2^{(i_2)}$, where $J_1^{(i_1)}$ denotes the i_1-th dyadic ancestor of J_1, similarly for $J_2^{(i_2)}$. Hence, the sum can be decomposed into three parts: $I_1 \subset J_1^{(i_1)}$ and $I_2 \subset J_2^{(i_2)}$ (regular), $I_1 \subset J_1^{(i_1)}$ and $I_2 \supsetneq J_2^{(i_2)}$, $I_1 \supsetneq J_1^{(i_1)}$ and $I_2 \subset J_2^{(i_2)}$ (mixed).

1) Regular case:

Following [7] one can decompose the arising sum into sums of classical bi-parameter dyadic paraproducts $B_0(b,f)$ and its slightly revised version $B_{k,l}(b,f)$: for any integers $k, l \geqslant 0$, $B_{k,l}$ is the bi-parameter dyadic paraproduct defined as

$$B_{k,l}(b,f) = \sum_{I,J} \beta_{IJ}\langle b, h_{I^{(k)}}\otimes u_{J^{(l)}}\rangle\langle f, h_I^{\epsilon_1}\otimes u_J^{\epsilon_2}\rangle h_I^{\epsilon_1'}\otimes u_J^{\epsilon_2'} |I^{(k)}|^{-1/2}|J^{(l)}|^{-1/2},$$

where β_{IJ} is a sequence satisfying $|\beta_{IJ}| \leqslant 1$. When $k > 0$, all Haar functions in the first variable are cancellative, while when $k = 0$, there is at most one of $h_I^{\epsilon_1}, h_I^{\epsilon_1'}$ being noncancellative. The same assumption goes for the second variable. Observe that when $k = l = 0$,

$B_{k,l}$ becomes the classical paraproduct B_0. It is proved in [7] that

$$\|B_{k,l}(b,f)\|_{L^2} \lesssim \|b\|_{\mathrm{BMO}}\|f\|_{L^2}$$

with a constant independent of k,l and the product BMO norm on the right hand side.

Then since little bmo functions are contained in product BMO, this part can be controlled. More specifically, write

$$\begin{aligned}
[b, S^{i_1j_1i_2j_2}]f &= \sum_{I_1,I_2}\sum_{J_1,J_2}\langle b, h_{I_1}\otimes h_{I_2}\rangle\langle f, h_{J_1}\otimes h_{J_2}\rangle h_{I_1}\otimes h_{I_2} S^{i_1j_1i_2j_2}(h_{J_1}\otimes h_{J_2})\\
&\quad - \sum_{I_1,I_2}\sum_{J_1,J_2}\langle b, h_{I_1}\otimes h_{I_2}\rangle\langle f, h_{J_1}\otimes h_{J_2}\rangle S^{i_1j_1i_2j_2}(h_{I_1}h_{J_1}\otimes h_{I_2}h_{J_2})\\
&=: I + II,
\end{aligned}$$

then one can estimate term I and II separately. According to the definition of dyadic shifts, term I is equal to

$$\begin{aligned}
&\sum_{J_1,J_2}\sum_{I_1:I_1\subset J_1^{(i_1)}}\sum_{I_2:I_2\subset J_2^{(i_2)}}\langle b, h_{I_1}\otimes h_{I_2}\rangle\langle f, h_{J_1}\otimes h_{J_2}\rangle h_{I_1}\otimes h_{I_2}\cdot\\
&\quad\left(\sum_{\substack{J_1':J_1'\subset J_1^{(i_1)}\\ \ell(J_1')=2^{i_1-j_1}\ell(J_1)}}\sum_{\substack{J_2':J_2'\subset J_2^{(i_2)}\\ \ell(J_2')=2^{i_2-j_2}\ell(J_2)}} a_{J_1J_1'J_1^{(i_1)}J_2J_2'J_2^{(i_2)}} h_{J_1'}\otimes h_{J_2'}\right)\\
&= \sum_{K_1,K_2}\sum_{J_1:J_1\subset K_1}^{(i_1)}\sum_{J_2:J_2\subset K_2}^{(i_2)}\sum_{I_1:I_1\subset K_1}\sum_{I_2:I_2\subset K_2}\langle b, h_{I_1}\otimes h_{I_2}\rangle\langle f, h_{J_1}\otimes h_{J_2}\rangle h_{I_1}\otimes h_{I_2}\cdot\\
&\quad\left(\sum_{J_1':J_1'\subset K_1}^{(j_1)}\sum_{J_2':J_2'\subset K_2}^{(j_2)} a_{J_1J_1'K_1J_2J_2'K_2} h_{J_1'}\otimes h_{J_2'}\right)\\
&= \sum_{I_1,I_2}\langle b, h_{I_1}\otimes h_{I_2}\rangle h_{I_1}\otimes h_{I_2}\sum_{\substack{K_1\supset I_1\\ K_2\supset I_2}}\\
&\quad\left(\sum_{J_1,J_1'\subset K_1}^{(i_1,j_1)}\sum_{J_2,J_2'\subset K_2}^{(i_2,j_2)} a_{J_1J_1'K_1J_2J_2'K_2}\langle f, h_{J_1}\otimes h_{J_2}\rangle h_{J_1'}\otimes h_{J_2'}\right)\\
&= \sum_{I_1,I_2}\langle b, h_{I_1}\otimes h_{I_2}\rangle h_{I_1}\otimes h_{I_2}\sum_{J_1':J_1'^{(j_1)}\supset I_1}\sum_{J_2':J_2'^{(j_2)}\supset I_2}\langle S^{i_1j_1i_2j_2}f, h_{J_1'}\otimes h_{J_2'}\rangle h_{J_1'}\otimes h_{J_2'}.
\end{aligned}$$

Because of the supports of Haar functions, the inner sum above can be further decomposed into four parts, where

$$I=\sum_{I_1,I_2}\sum_{J_1'\supsetneq I_1}\sum_{J_2'\supsetneq I_2},\quad II=\sum_{I_1,I_2}\sum_{J_1'\supsetneq I_1}\sum_{J_2':J_2'\subset I_2\subset J_2'^{(j_2)}}$$
$$III=\sum_{I_1,I_2}\sum_{J_1':J_1'\subset I_1\subset J_1'^{(j_1)}}\sum_{J_2'\supsetneq I_2},\quad IV=\sum_{I_1,I_2}\sum_{J_1':J_1'\subset I_1\subset J_1'^{(j_1)}}\sum_{J_2':J_2'\subset I_2\subset J_2'^{(j_2)}}.$$

Hence, similarly as in [7], one has

$$I=\sum_{I_1,I_2}\langle b,h_{I_1}\otimes h_{I_2}\rangle\langle S^{i_1j_1i_2j_2}f,h^1_{J_1'}\otimes h^1_{J_2'}\rangle h_{I_1}\otimes h_{I_2}|I_1|^{-1/2}|I_2|^{-1/2},$$

which is a bi-parameter paraproduct $B_0(b,f)$. Moreover, one has

$$\begin{aligned}II&=\sum_{I_1,I_2}\langle b,h_{I_1}\otimes h_{I_2}\rangle h_{I_1}\otimes h_{I_2}\sum_{J_2':J_2'\subset I_2\subset J_2'^{(j_2)}}\langle S^{i_1j_1i_2j_2}f,h^1_{I_1}\otimes h_{J_2'}\rangle|I_1|^{-1/2}h_{J_2'}\\&=\sum_{l=0}^{j_2}\sum_{I_1,J_2'}\beta_{J_2'}\langle b,h_{I_1}\otimes h_{J_2'^{(l)}}\rangle\langle S^{i_1j_1i_2j_2}f,h^1_{I_1}\otimes h_{J_2'}\rangle h_{I_1}\otimes h_{J_2'}|I_1|^{-1/2}|J_2'^{(l)}|^{-1/2}\\&=\sum_{l=0}^{j_2}B_{0l}(b,S^{i_1j_1i_2j_2}f),\end{aligned}$$

where constants $\beta_{J_2'}\in\{1,-1\}$, and B_{0l} are the generalized bi-parameter paraproducts of type $(0,l)$ defined in [7] whose $L^2\to L^2$ operator norm is uniformly bounded by $\|b\|_{\mathrm{BMO}}$ product BMO. Similarly, one can show that

$$III=\sum_{k=0}^{j_1}B_{k0}(b,S^{i_1j_1i_2j_2}f),\quad IV=\sum_{k=0}^{j_1}\sum_{l=0}^{j_2}B_{kl}(b,S^{i_1j_1i_2j_2}f).$$

Since $\|b\|_{\mathrm{BMO}}\lesssim\|b\|_{\mathrm{bmo}}$, all the forms above are L^2 bounded. This completes the discussion of term I.

To get an estimate of term II, we need to decompose it into finite linear combinations of $S^{i_1j_1i_2j_2}(B_{kl}(b,f))$. By linearity, one can write $S^{i_1j_1i_2j_2}$ on the outside from the beginning, and we will only look at the inside sum. Recall that in [7], one splits for example the sum regarding the first variable into three parts: $I_1\subsetneq J_1$, $I_1=J_1$, $J_1\subsetneq I_1\subset J_1^{(i_1)}$. If we split the second variable as well, there are nine mixed parts, and it's not hard to show that each of them can be represented as a finite sum of $B_{kl}(b,f)$. We omit the details.

2) Mixed case. Let's call the second and the third 'mixed' parts, and as the two are symmetric, it suffices to look at the second one, i.e. $I_1\subset J_1^{(i_1)},I_2\supsetneq J_2^{(i_2)}$. In the first variable, we still have the old case $I_1\subset J_1^{(i_1)}$ that appeared in [7], so morally speaking, we only need to nicely play around with the stronger little bmo norm to handle the second variable. For any fixed I_1, J_1, I_2, J_2, since $I_2 \supsetneq J_2^{(i_2)}$, the definition of dyadic shifts implies that

$$h_{I_1} \otimes h_{I_2} S^{i_1 j_1 i_2 j_2}(h_{J_1} \otimes h_{J_2}) = h_{I_1} S^{i_1 j_1 i_2 j_2}(h_{J_1} \otimes h_{I_2} h_{J_2})$$

and

$$S^{i_1 j_1 i_2 j_2}(h_{i_1} h_{J_1} \otimes h_{I_2} h_{J_2}) = h_{I_2} S^{i_1 j_1 i_2 j_2}(h_{I_1} h_{J_1} \otimes h_{J_2}).$$

Hence, we still have cancellation in the second variable, which converts the mixed case to

$$\begin{aligned}
&\sum_{I_1 \subset J_1^{(i_1)}} \sum_{I_2 \supsetneq J_2^{(i_2)}} \langle b, h_{I_1} \otimes h_{I_2} \rangle \langle f, h_{J_1} \otimes h_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}](h_{J_1} \otimes h_{I_2} h_{J_2}) \\
&= \sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle f, h_{J_1} \otimes h_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}](h_{J_1} \otimes \sum_{I_2 \supsetneq J_2^{(i_2)}} \langle b, h_{I_1} \otimes h_{I_2} \rangle h_{I_2} h_{J_2}) \\
&= \sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle f, h_{J_1} \otimes h_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}](h_{J_1} \otimes \langle b, h_{I_1} \otimes h^1_{J_2^{(i_2)}} \rangle h^1_{J_2^{(i_2)}} h_{J_2}) \\
&= \sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle b, h_{I_1} \otimes h^1_{J_2^{(i_2)}} \rangle |J_2^{(i_2)}|^{-1/2} \langle f, h_{J_1} \otimes h_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}](h_{J_1} \otimes h_{J_2}) \\
&= \sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{I_1} \rangle_1 \langle f, h_{J_1} \otimes h_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}](h_{J_1} \otimes h_{J_2}),
\end{aligned}$$

where $\langle b \rangle_{J_2^{(i_2)}}$ denotes the average value of b on $J_2^{(i_2)}$, which is a function of only the first variable.

In the following, we will once again estimate the first term and second term of the commutator separately, and the L^2 norm of each of them will be proved to be bounded by $\|b\|_{\text{bmo}} \|f\|_{L^2}$.

a) First term.

By definition of the dyadic shift, the first term is equal to

$$\sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{I_1} \rangle_1 h_{I_1} \langle f, h_{J_1} \otimes h_{J_2} \rangle \cdot$$

$$\left(\sum_{\substack{J_1' \subset J_1^{(i_1)} \\ \ell(J_1') = 2^{i_1 - j_1} \ell(J_1)}} \sum_{\substack{J_2' \subset J_2^{(i_2)} \\ \ell(J_2') = 2^{i_2 - j_2} \ell(J_2)}} a_{J_1 J_1' J_1^{(i_1)} J_2 J_2' J_2^{(i_2)}} h_{J_1'} \otimes h_{J_2'} \right),$$

which by reindexing $K_1 := J_1^{(i_1)}$ is the same as

$$\sum_{I_1,J_2}\langle\langle b\rangle_{J_2^{(i_2)}},h_{I_1}\rangle_1 h_{I_1}\sum_{K_1:K_1\supset I_1}\sum_{J_1\subset K_1}^{(i_1)}\sum_{J_1'\subset K_1}^{(j_1)}\sum_{J_2'\subset J_2^{(i_2)}}^{(j_2)} a_{J_1J_1'K_1J_2J_2'J_2^{(i_2)}}\langle f,h_{J_1}\otimes h_{J_2}\rangle h_{J_1'}\otimes h_{J_2'}$$
$$=\sum_{I_1,J_2}\langle\langle b\rangle_{J_2^{(i_2)}},h_{I_1}\rangle_1 h_{I_1}\sum_{J_1':J_1'^{(j_1)}\supset I_1} h_{J_1'}\otimes\langle S^{i_1j_1i_2j_2}(\langle f,h_{J_2}\rangle_2\otimes h_{J_2}),h_{J_1'}\rangle_1,$$

where the inner sum is the orthogonal projection of the image of $\langle f,h_{J_2}\rangle_2\otimes h_{J_2}$ under $S^{i_1j_1i_2j_2}$ onto the span of $\{h_{J_1'}\}$ such that $J_1'^{(j_1)}\supset I_1$. Taking into account of the supports of the Haar functions in the first variable, one can further split the sum into two parts where

$$I:=\sum_{J_2}\sum_{I_1\subsetneq J_1'},\quad II:=\sum_{J_2}\sum_{J_1'\subset I_1\subset J_1'^{(j_1)}}.$$

Summing over J_1' first implies that

$$I=\sum_{J_2}\sum_{I_1}\langle\langle b\rangle_{J_2^{(i_2)}},h_{I_1}\rangle_1 h_{I_1}\left(h^1_{I_1}\otimes\langle S^{i_1j_1i_2j_2}(\langle f,h_{J_2}\rangle_2\otimes h_{J_2}),h^1_{I_1}\rangle_1\right)$$
$$=:\sum_{J_2}B_0(\langle b\rangle_{J_2^{(i_2)}},S^{i_1j_1i_2j_2}(\langle f,h_{J_2}\rangle_2\otimes h_{J_2}))$$

where $B_0(b,f):=\sum_I\langle b,h_I\rangle\langle f,h_I^1\rangle h_I|I|^{-1/2}$ is a classical one-parameter paraproduct in the first variable. Note that the L^2 norm of it is bounded by $\|b\|_{\mathrm{BMO}}\|f\|_{L^2}$. Moreover, according to the definition of $S^{i_1j_1i_2j_2}$, for any fixed J_2

$$S^{i_1j_1i_2j_2}(\langle f,h_{J_2}\rangle_2\otimes h_{J_2})=\sum_{J_2':J_2'^{(j_2)}=J_2^{(i_2)}}\langle S^{i_1j_1i_2j_2}(\langle f,h_{J_2}\rangle_2\otimes h_{J_2}),h_{J_2'}\rangle_2\otimes h_{J_2'}.$$

In other words, $S^{i_1j_1i_2j_2}(\langle f,h_{J_2}\rangle_2\otimes h_{J_2})$ only lives on the span of $\{h_{J_2'}:J_2'^{(j_2)}=J_2^{(i_2)}\}$. Hence, by linearity there holds

$$I=\sum_{J_2}\sum_{J_2':J_2'^{(j_2)}=J_2^{(i_2)}}B_0\big(\langle b\rangle_{J_2^{(i_2)}},\langle S^{i_1j_1i_2j_2}(\langle f,h_{J_2}\rangle_2\otimes h_{J_2}),h_{J_2'}\rangle_2\big)\otimes h_{J_2'}$$
$$=\sum_{J_2'}\left(B_0\big(\langle b\rangle_{J_2'^{(j_2)}},\langle S^{i_1j_1i_2j_2}(\sum_{J_2:J_2^{(i_2)}=J_2'^{(j_2)}}\langle f,h_{J_2}\rangle_2\otimes h_{J_2}),h_{J_2'}\rangle_2\big)\right)\otimes h_{J_2'}.$$

Thus, orthogonality in the second variable implies that

$$\|I\|^2_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}$$
$$=\sum_{J_2'}\|B_0\big(\langle b\rangle_{J_2'^{(j_2)}},\langle S^{i_1j_1i_2j_2}(\sum_{J_2:J_2^{(i_2)}=J_2'^{(j_2)}}\langle f,h_{J_2}\rangle_2\otimes h_{J_2}),h_{J_2'}\rangle_2\big)\|^2_{L^2(\mathbb{R}^n)}$$
$$\lesssim\sum_{J_2'}\|\langle b\rangle_{J_2'^{(j_2)}}\|^2_{\mathrm{BMO}(\mathbb{R}^n)}\|\langle S^{i_1j_1i_2j_2}(\sum_{J_2:J_2^{(i_2)}=J_2'^{(j_2)}}\langle f,h_{J_2}\rangle_2\otimes h_{J_2}),h_{J_2'}\rangle_2\|^2_{L^2(\mathbb{R}^n)}.$$

Observing that $\|\langle b\rangle_{J_2'^{(j_2)}}\|_{\mathrm{BMO}(\mathbb{R}^n)} \leqslant \langle \|b\|_{\mathrm{BMO}(\mathbb{R}^n)}\rangle_{J_2'^{(j_2)}} \leqslant \|b\|_{\mathrm{bmo}}$, one has

$$\leqslant \|b\|_{\mathrm{bmo}}^2 \sum_{J_2'} \|\langle S^{i_1j_1i_2j_2}(\sum_{J_2 : J_2^{(i_2)} = J_2'^{(j_2)}} \langle f, h_{J_2}\rangle_2 \otimes h_{J_2}), h_{J_2'}\rangle_2\|_{L^2(\mathbb{R}^n)}^2$$

$$= \|b\|_{\mathrm{bmo}}^2 \| \sum_{J_2'} \langle S^{i_1j_1i_2j_2}(\sum_{J_2 : J_2^{(i_2)} = J_2'^{(j_2)}} \langle f, h_{J_2}\rangle_2 \otimes h_{J_2}), h_{J_2'}\rangle_2 \otimes h_{J_2'}\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}^2.$$

Note that the sum in the L^2 norm is in fact very simple:

$$\begin{aligned}
&\sum_{J_2'} \langle S^{i_1j_1i_2j_2}(\sum_{J_2 : J_2^{(i_2)} = J_2'^{(j_2)}} \langle f, h_{J_2}\rangle_2 \otimes h_{J_2}), h_{J_2'}\rangle_2 \otimes h_{J_2'} \\
&= \sum_{J_2} \sum_{J_2' : J_2'^{(j_2)} = J_2^{(i_2)}} \langle S^{i_1j_1i_2j_2}(\langle f, h_{J_2}\rangle_2 \otimes h_{J_2}), h_{J_2'}\rangle_2 \otimes h_{J_2'} \\
&= \sum_{J_2} S^{i_1j_1i_2j_2}(\langle f, h_{J_2}\rangle_2 \otimes h_{J_2}) = S^{i_1j_1i_2j_2}(f).
\end{aligned}$$

Hence, the uniform boundedness of the $L^2 \to L^2$ operator norm of dyadic shifts implies that

$$\|I\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}^2 \lesssim \|b\|_{\mathrm{bmo}}^2 \|f\|_{L^2(\mathbb{R}^n)\times\mathbb{R}^m}^2.$$

In order to handle II, we split it into a finite sum depending on the levels of I_1 upon J_1', which leads to

$$\begin{aligned}
II &= \sum_{k=0}^{j_1} \sum_{J_2} \sum_{J_1'} \langle \langle b\rangle_{J_2^{(i_2)}}, h_{J_1'^{(k)}}\rangle_1 h_{J_1'^{(k)}} h_{J_1'} \otimes \langle S^{i_1j_1i_2j_2}(\langle f, h_{J_2}\rangle_2 \otimes h_{J_2}), h_{J_1'}\rangle_1 \\
&= \sum_{k=0}^{j_1} \sum_{J_2} \sum_{J_1'} \beta_{J_1',k} |J_1'^{(k)}|^{-1/2} \langle \langle b\rangle_{J_2^{(i_2)}}, h_{J_1'^{(k)}}\rangle_1 h_{J_1'} \otimes \langle S^{i_1j_1i_2j_2}(\langle f, h_{J_2}\rangle_2 \otimes h_{J_2}), h_{J_1'}\rangle_1 \\
&=: \sum_{k=0}^{j_1} \sum_{J_2} B_k(\langle b\rangle_{J_2^{(i_2)}}, S^{i_1j_1i_2j_2}(\langle f, h_{J_2}\rangle_2 \otimes h_{J_2})),
\end{aligned}$$

where $B_k(b, f) := \sum_I \beta_{I,k} \langle b, h_{I^{(k)}}\rangle \langle f, h_I\rangle h_I |I^{(k)}|^{-1/2}$ is a generalized one-parameter paraproduct studied in [7], whose L^2 norm is uniformly bounded by $\|b\|_{\mathrm{BMO}}\|f\|_{L^2}$, independent of k and the coefficients $\beta_{I,k} \in \{1, -1\}$. Then one can proceed as how we dealt with part I to conclude that

$$\|II\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)} \lesssim (1 + j_1)\|b\|_{\mathrm{bmo}}\|f\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)},$$

which together with the estimate for part I implies that

$$\|\text{First term}\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)} \lesssim (1 + j_1)\|b\|_{\mathrm{bmo}}\|f\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}.$$

b) Second term.

As the second term by linearity is the same as

$$S^{i_1 j_1 i_2 j_2}\left(\sum_{J_2}\sum_{I_1\subset J_1^{(i_1)}}\langle\langle b\rangle_{J_2^{(i_2)}}, h_{I_1}\rangle_1\langle f, h_{J_1}\otimes h_{J_2}\rangle h_{I_1}h_{J_1}\otimes h_{J_2}\right),$$

the $L^2\to L^2$ boundedness of the shift implies that it suffices to estimate the L^2 norm of the term inside the parentheses. Since $I_1\cap J_1\neq\varnothing$, one can further split the sum into two parts:

$$I := \sum_{J_2}\sum_{I_1\subsetneq J_1}, \quad II := \sum_{J_2}\sum_{J_1\subset I_1\subset J_1^{(i_1)}}.$$

Summing over J_1 first implies that

$$\begin{aligned} I &= \sum_{J_2}\sum_{I_1}\langle\langle b\rangle_{J_2^{(i_2)}}, h_{I_1}\rangle_1\langle f, h^1_{I_1}\otimes h_{J_2}\rangle h_{I_1}h^1_{I_1}\otimes h_{J_2} \\ &=: \sum_{J_2} B_0(\langle b\rangle_{J_2^{(i_2)}}, \langle f, h_{J_2}\rangle_2)\otimes h_{J_2}, \end{aligned}$$

where $B_0(b,f) := \sum_I\langle b, h_I\rangle\langle f, h_I^1\rangle h_I|I|^{-1/2}$ is a classical one-parameter paraproduct in the first variable. Hence,

$$\begin{aligned} \|I\|^2_{L^2(\mathbb{R}^n\times\mathbb{R}^m)} &= \sum_{J_2}\|B_0(\langle b\rangle_{J_2^{(i_2)}}, \langle f, h_{J_2}\rangle_2)\|^2_{L^2(\mathbb{R}^n)} \\ &\lesssim \sum_{J_2}\|\langle b\rangle_{J_2^{(i_2)}}\|^2_{\mathrm{BMO}(\mathbb{R}^n)}\|\langle f, h_{J_2}\rangle_2\|^2_{L^2(\mathbb{R}^n)} \\ &\leqslant \|b\|^2_{\mathrm{bmo}}\sum_{J_2}\|\langle f, h_{J_2}\rangle_2\|^2_{L^2(\mathbb{R}^n)} = \|b\|^2_{\mathrm{bmo}}\|f\|^2_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}. \end{aligned}$$

For part II, note that it can be decomposed as

$$\begin{aligned} II &= \sum_{k=0}^{i_1}\sum_{J_2}\sum_{J_1}\langle\langle b\rangle_{J_2^{(i_2)}}, h_{J_1^{(k)}}\rangle_1\langle f, h_{J_1}\otimes h_{J_2}\rangle h_{J_1^{(k)}}h_{J_1}\otimes h_{J_2} \\ &= \sum_{k=0}^{i_1}\sum_{J_2}\sum_{J_1}\beta_{J_1,k}|J_1^{(k)}|^{-1/2}\langle\langle b\rangle_{J_2^{(i_2)}}, h_{J_1^{(k)}}\rangle_1\langle\langle f, h_{J_2}\rangle_2, h_{J_1}\rangle_1 h_{J_1}\otimes h_{J_2} \\ &=: \sum_{k=0}^{i_1}\sum_{J_2} B_k(\langle b\rangle_{J_2^{(i_2)}}, \langle f, h_{J_2}\rangle_2)\otimes h_{J_2}, \end{aligned}$$

where coefficients $\beta_{J_1,k}\in\{1,-1\}$ and the L^2 norm of the generalized paraproduct B_k is uniformly bounded as mentioned before. Therefore, a same argument as for part I shows that

$$\|II\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)} \lesssim (1+i_1)\|b\|_{\mathrm{bmo}}\|f\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)},$$

which completes the discussion of the second term, and thus proves that the mixed case is bounded. □

The upper bound result we just proved can be extended to $\mathbb{R}^{\vec{d}}$, to arbitrarily many parameters and an arbitrary number of iterates in the commutator. To do this, consider multi-parameter singular integral operators studied in [26], which satisfy a weak boundedness property and are paraproduct free, meaning that any partial adjoint of T is zero if acting on some tensor product of functions with one of the components being 1. And consider a little product BMO function $b \in \mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$. One can then prove

Theorem 4.4. *Let us consider $\mathbb{R}^{\vec{d}}$, $\vec{d} = (d_1, \dots, d_t)$ with a partition $\mathcal{I} = (I_s)_{1 \leqslant s \leqslant l}$ of $\{1, \dots, t\}$ as discussed before. Let $b \in BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ and let T_s denote a multi-parameter paraproduct free Journé operator acting on function defined on $\bigotimes_{k \in I_s} \mathbb{R}^{d_k}$. Then we have the estimate below*

$$\|[T_1, \dots [T_l, b] \dots]\|_{L^2(\mathbb{R}^{\vec{d}}) \circlearrowleft} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

The part of the proof that targets the Journé operators proceeds exactly the same as the bi-parameter case with the multi-parameter version of the representation theorem proven in [26]. Certainly, as the number of parameters increases, more mixed cases will appear. However, if one follows the corresponding argument above for each variable in each case, it is not hard to check that eventually, the boundedness of the arising paraproducts is implied exactly by the little product BMO norm of the symbol. The difficulty of higher iterates is overcome in observing that the commutator splits into commutators with no iterates, as was done in [7]. We omit the details.

The assumption that the operators be paraproduct free is sufficient for us. The general case is currently under investigation by one of the authors. Important to our arguments for lower bounds with Riesz transforms is the corollary below, which follows from the control on the norm of the estimate in Theorem 4.4 by an application of triangle inequality. It is a stability result for characterising families of Journé operators.

Corollary 4.5. *Let for every $1 \leqslant s \leqslant l$ be given a collection $\mathcal{T}_s = \{T_{s,j_s}\}$ of paraproduct free Journé operators on $\bigotimes_{k \in I_s} \mathbb{R}^{d_k}$ that characterise $BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ via a two-sided commutator estimate*

$$\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{j}} \|[T_{1,j_1}, \dots [T_{l,j_l}, b] \dots]\|_{L^2(\mathbb{R}^{\vec{d}}) \circlearrowleft} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

Then there exists $\varepsilon > 0$ such that for any family of paraproduct free Journé operators $\mathcal{T}'_s = \{T'_{s,j_s}\}$ with characterising constants $\|T'_{s,j_s}\|_{CZ} \leqslant \varepsilon$, the family $\{T_{s,j_s} + T'_{s,j_s}\}$ still characterises $BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$.

5. Real variables: lower bounds

In this section, we are again in $\mathbb{R}^{\vec{d}}$. It is our aim to prove the following characterisation theorem of the space $\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$.

Theorem 5.1. *The following are equivalent with linear dependence in the respective norms.*

(1) $b \in BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$

(2) All commutators of the form $[R_{k_1,j_{k_1}}, \ldots, [R_{k_l,j_{k_l}}, b] \ldots]$ *are bounded in* $L^2(\mathbb{R}^{\vec{d}})$ *where* $k_s \in I_s$ *and* $R_{k_s,j_{k_s}}$ *is the one-parameter Riesz transform in direction* j_{k_s}.

(3) All commutators of the form $[\vec{R}_{1,\vec{j}^{(1)}}, \ldots, [\vec{R}_{l,\vec{j}^{(l)}}, b] \ldots]$ *are bounded in* $L^2(\mathbb{R}^{\vec{d}})$ *where* $\vec{j}^{(s)} = (j_k)_{k \in I_s}$, $1 \leqslant j_k \leqslant d_k$ *and the operators* $\vec{R}_{s,\vec{j}^{(s)}}$ *are a tensor product of Riesz transforms* $\vec{R}_{s,\vec{j}^{(s)}} = \bigotimes_{k \in I_s} R_{k,j_k}$.

Corollary 5.2. *Let* $\vec{j} = (j_1, \ldots, j_t)$ *with* $1 \leqslant j_k \leqslant d_k$ *and let for each* $1 \leqslant s \leqslant l$, $\vec{j}^{(s)} = (j_k)_{k \in I_s}$ *be associated a tensor product of Riesz transforms* $\vec{R}_{s,\vec{j}^{(s)}} = \bigotimes_{k \in I_s} R_{k,j_k}$*; here* R_{k,j_k} *are* j_k^{th} *Riesz transforms acting on functions defined on the* k^{th} *variable. We have the two-sided estimate*

$$\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{j}} \|[\vec{R}_{1,\vec{j}^{(1)}}, \ldots, [\vec{R}_{t,\vec{j}^{(t)}}, b] \ldots]\|_{L^2(\mathbb{R}^{\vec{d}}) \circlearrowleft} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

Such two-sided estimates also hold in L^p for $1 < p < \infty$. Remarks will be made in section 7. From the inductive nature of our arguments, it will also be apparent that the characterisation holds when we consider intermediate cases, meaning commutators with any fixed number of Riesz transforms in each iterate. Below we state our most general two-sided estimate through Riesz transforms.

Theorem 5.3. *Let* $1 < p < \infty$. *Under the same assumptions as Corollary 5.2 and for any fixed* $\vec{n} = (n_s)$ *where* $1 \leqslant n_s \leqslant |I_s|$, *we have the two-sided estimate*

$$\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{j}} \|[\vec{R}_{1,\vec{j}^{(1)}}, \ldots, [\vec{R}_{l,\vec{j}^{(l)}}, b] \ldots]\|_{L^p(\mathbb{R}^{\vec{d}}) \circlearrowleft} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}$$

where $\vec{j}^{(s)} = (j_k)_{k \in I_s}$, $0 \leqslant j_k \leqslant d_k$ *and for each* s, *there are* n_s *non-zero choices. A Riesz transform in direction 0 is understood as the identity.*

For $p = 2$ and $\vec{n} = \vec{1}$ this is the equivalence *(1)*$\Leftrightarrow$ *(2)* and for $\vec{n} = (|I_1|, \ldots, |I_l|)$ it is the equivalence *(1)*$\Leftrightarrow$ *(3)* from Theorem 5.1.

The statements above also serve as the statement of the general case for products of Hilbert transforms. Infact, when any $d_k = 1$ just replace the Riesz transforms by the Hilbert transform in that variable. In this section, we consider the case $d_k \geqslant 2$ for $1 \leqslant k \leqslant t$ and thus iterated commutators with tensor products of Riesz transforms only. The special case when $d_k = 1$ for some k is easier but requires extra care for notation, which is why we omit it here.

The proof in the Hilbert transform case relied heavily on analytic projections and orthogonal spaces, a feature that we do not have when working with Riesz transforms. We are going to simulate the one-dimensional case by a two-step passage via intermediary Calderón-Zygmund operators whose multiplier symbols are adapted to cones.

In dimension $d \geqslant 2$, a cone $C \subset \mathbb{R}^d$ with cubic base is given by the data (ξ, Q) where $\xi \in \mathbb{S}^{d-1}$ is the direction of the cone and the cube $Q \subset \xi^{\perp}$ centered at the origin is its aperture. The cone consists of all vectors θ that take the form $(\theta_\xi \xi, \theta_\perp)$ where $\theta_\xi = \langle \theta, \xi \rangle$ and $\theta_\perp \in \theta_\xi Q$. By λC we mean the dilated cone with data $(\xi, \lambda Q)$.

A cone D with ball base has data (ξ, r) for $0 < r < \pi/2$ and $\xi \in \mathbb{S}^{d-1}$ and consists of the vectors $\{\eta \in \mathbb{R}^d : d(\xi, \eta/\|\eta\|) \leqslant r\}$ where d is the geodesic distance (with distance of antipodal points being π.)

Given any cone C or D, we consider its Fourier projection operator defined via $\widehat{P_C f} = \mathbf{1}_C \hat{f}$. When the apertures are cubes, such operators are combinations of Fourier projections onto half spaces and as such admit uniform L^p bounds. Among others, this fact made cubic cones necessary in the considerations in [17] and [8] that we are going to need. For further technical reasons in the proof of an error term in an underlying lower estimate in [17] and [8], these operators are not quite good enough, mainly because they are not of Calderón-Zygmund type. For a given cone C, consider a Calderón-Zygmund operator T_C with a kernel K_C whose Fourier symbol $\widehat{K_C} \in C^\infty$ and satisfies the estimate $\mathbf{1}_C \leqslant \widehat{K_C} \leqslant \mathbf{1}_{(1+\tau)C}$. This is accomplished by mollifying the symbol $\mathbf{1}_C$ of the cone projection associated to cone C on $\mathbb{S}^{d-1}$ and then extending radially. We use the same definition for T_D.

Given a collection of cones $\vec{C} = (C_k)$ we denote by $T_{\vec{C}}, P_{\vec{C}}$ the corresponding tensor product operators.

In [17] it has been proven that Calderón-Zygmund operators adapted to certain cones of cubic aperture classify product BMO via commutators. As part of the argument, it was observed that test functions with opposing Fourier supports made the commutator large. This observation reduced the terms arising in the commutator to those resembling Hankel operators. In [8] a refinement was proven, that will be helpful to us. We prefer to work with cones with round base. Lower bounds for such commutators can be deduced from the assertion of the main theorem in [8], but we need to preserve the information on the Fourier support of the test function that makes the commutators large in order to succeed with our argument. Information on this test function replaces the property that the Hilbert transform is a combination of orthogonal projections of two orthogonal half spaces. It is instrumental to our argument. We have the following lemma, very similar to that in [17] and [8], the only difference being that the cones are based on balls instead of cubes.

Lemma 5.4. *For every parameter $1 \leqslant k \leqslant t$ there exist a finite set of directions $\Upsilon_k \in \mathbb{S}^{d_k-1}$ and an aperture $0 < r_k < \pi/2$ so that for every b, there exist cones $D_k = D(\xi_k, r_k)$ with $\xi_k \in \Upsilon_k$ as well as a normalized test function $f = \bigotimes_{k=1}^t f_k$ whose components have Fourier support in the opposing cones $D(-\xi_k, r_k)$ so that*

$$\|[T_{1,D_1} \ldots, [T_{t,D_t}, b] \ldots] f\|_2 \gtrsim \|b\|_{BMO_{(1)\ldots(t)}(\mathbb{R}^{\vec{d}})}.$$

The stress is on the fact that the collection is finite, somewhat specific and serves all admissible product BMO functions.

Proof. The lemma in [8] supplies us with the sets of directions Υ_k as well as cones of cubic aperture Q_k and a test function f supported in the opposing cones. Now choose the aperture r_k large enough so that $(1+\tau)C(\xi_k, Q_k) \subset D(\xi_k, r_k)$. Then we have the commutator estimate

$$\|[T_{1,D_1}..., [T_{t,D_t}, b]...]f\|_2 \gtrsim \|b\|_{\mathrm{BMO}_{(1)...(t)}(\mathbb{R}^{\vec{d}})}.$$

In fact, both commutators with cones C and D are L^2 bounded and reduce to $\|T_{\vec{D}}(bf)\|_2$ or $\|T_{\vec{C}}(bf)\|_2$ respectively thanks to the opposing Fourier support of f. Observe that $T_{\vec{C}}(bf) = T_{\vec{D}}(T_{\vec{C}}(bf)) = T_{\vec{C}}(T_{\vec{D}}(bf))$. With $\|T_{\vec{C}}\|_{2\to 2} \leqslant 1$, we see that $\|T_{\vec{D}}(bf)\|_2 \geqslant \|T_{\vec{C}}(bf)\|_2$. □

Using this a priori lower estimate, we are going to prove the lemma below.

Lemma 5.5. *Let us suppose we are in $\mathbb{R}^{\vec{d}}$ with a partition $\mathcal{I} = (I_s)$. For every $1 \leqslant k \leqslant t$ there exists a finite set of directions $\Upsilon_k \subset \mathbb{S}^{d_k-1}$ and an aperture r_k so that the following hold for all $b \in BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$:*

(1) For every $1 \leqslant s \leqslant l$ there exists a coordinate $v_s \in I_s$ and a direction $\xi_{v_s} \in \Upsilon_{v_s}$ and so that with the choice of cone $D_{v_s} = D(\xi_{v_s}, r_{v_s})$ and arbitrary D_k for coordinates $k \in I_s \backslash \{v_s\}$ and their tensor product $\vec{D}_s$ we have

$$\|[T_{1,\vec{D}_1}, \ldots, [T_{l,\vec{D}_l}, b] \ldots]\|_{2\to 2} \gtrsim \|b\|_{BMO_{\mathcal{I}}},$$

(2) The test function $f = \bigotimes_{k=1}^t$ which gives us a large L^2 norm in (1) has Fourier supports of the f_k contained in $D(-\xi_k, r_k)$ when $k = v_s$ and in D_k else.

Before we can begin with the proof of Lemma 5.5, we will need a real variable version of the facts on Toeplitz operators used earlier.

Lemma 5.6. *Let D_k for $1 \leqslant k \leqslant t$ denote any cones with respect to the k^{th} variable. Let T_{D_k} denote the adapted Calderón-Zygmund operators. Let K be any proper subset of $\{k : 1 \leqslant k \leqslant t\}$, let $\vec{D}_K = \bigotimes_{k\in K} D_k$ and $T_{\vec{D}_K}$ the associated tensor product of Calderón-Zygmund operators. Let $P^{\sigma}_{\vec{D}_K}$ be a tensor product of projection operators on cones $D(\xi_k, r_k)$ or opposing cones $D(-\xi_k, r_k)$. Let $j \notin K$. Then*

$$\|T_{\vec{D}_K} T_{D_j} b P^{\sigma}_{\vec{D}_K} P_{D_j}\|_{L^2(\mathbb{R}^{\vec{d}})\circlearrowleft} = \|T_{\vec{D}_K} b P^{\sigma}_{\vec{D}_K}\|_{L^2(\mathbb{R}^{\vec{d}})\circlearrowleft}.$$

Proof. As in the Hilbert transform case, we will establish this by composing some unilateral shift operators and studying their Fourier transform in the j variable. Let ξ_j denote the direction of the cone D_j, for any l define the shift operator

$$S_l g(x_j) = \int_{\mathbb{R}^{d_j}} \hat{g}(\eta_j) e^{2\pi i (l\xi_j + \eta_j)x_j}\, d\eta_j.$$

S_l is a translation operator on the Fourier side along the direction ξ_j of the cone D_j. It is not hard to observe that $S_l^* = S_{-l}$. Now define

$$A_l = S_{-l}T_{\vec{D}_K}T_{D_j}bP^{\sigma}_{\vec{D}_K}P_{D_j}S_l, \text{ and } B = T_{\vec{D}_K}bP^{\sigma}_{\vec{D}_K}.$$

We will prove that as $l \to +\infty$, $A_l \to B$ in the strong operator topology. As in the argument in Lemma 3.6, this together with the fact that S_l is an isometry will complete the proof. To see the convergence, let's first remember that S_l only acts on the j variable, and one always has the identities

$$S_lS_{-l} = Id \quad \text{and} \quad S_{-l}bS_l = b.$$

This implies

$$\begin{aligned} A_l &= T_{\vec{D}_K}(S_{-l}T_{D_j}S_l)(S_{-l}bS_l)P^{\sigma}_{\vec{D}_K}(S_{-l}P_{D_j}S_l) \\ &= T_{\vec{D}_K}(S_{-l}T_{D_j}S_l)bP^{\sigma}_{\vec{D}_K}(S_{-l}P_{D_j}S_l). \end{aligned}$$

We claim that both $S_{-l}T_{D_j}S_l$ and $S_{-l}P_{D_j}S_l$ converge to the identity operator in the strong operator topology, which then imply that $A_l \to B$ as $l \to \infty$. Next we will only prove $S_{-l}T_{D_j}S_l \to Id$ as the second limit is almost identical. Observe that $\|S_{-l}T_{D_j}S_lf - f\| = \|(T_{D_j} - I)S_lf\|$. Given any L^2 function f and any fixed large $l \geqslant 0$. Consider the f with frequencies supported in $\mathbb{R}^{d_1} \times \ldots \times (D_j - l\xi_j) \times \ldots \times \mathbb{R}^{d_t}$. In this case, S_lf has Fourier support in $\mathbb{R}^{d_1} \times \ldots \times D_j \times \ldots \times \mathbb{R}^{d_t}$ where the symbol of T_{D_j} equals 1. Thus, for such f, we have $S_{-l}T_{D_j}S_lf = f$. The sets $\mathbb{R}^{d_1} \times \ldots \times (D_j - l\xi_j) \times \ldots \times \mathbb{R}^{d_t}$ exhaust the frequency space. With $\|T_{D_j} - I\|_{2\to 2} \leqslant 1$ the operators $S_{-l}T_{D_j}S_l$ converge to the Identity in the strong operator topology, and the lemma is proved. Observe that the aperture of the cone D_j is not relevant to the proof. □

We proceed with the proof of the lower estimate for cone transforms.

Proof. (of Lemma 5.5) For a given symbol $b \in \mathrm{BMO}_{\mathcal{I}}$, there exist for all $1 \leqslant s \leqslant l$ coordinates $\boldsymbol{v} = (v_s), v_s \in I_s$ and a choice of variables not indexed by v_s, $\vec{x}^{\,0}_{\hat{\boldsymbol{v}}}$ so that up to an arbitrarily small error

$$\|b\|_{\mathrm{BMO}_{\mathcal{I}}} = \|b(\vec{x}^{\,0}_{\hat{\boldsymbol{v}}})\|_{\mathrm{BMO}_{(v_1)\ldots(v_l)}}.$$

By Lemma 5.4, there exist cones $D_{v_s} = D(\xi_{v_s}, r_{v_s})$ with directions $\xi_{v_s} \in \Upsilon_{v_s}$ and a normalized test function f_H in variables v_s with opposing Fourier support such that we have the lower estimate

$$\|[T_{v_1,D_{v_1}}, \ldots, [T_{v_l,D_{v_l}}, b(\vec{x}^{\,0}_{\hat{\boldsymbol{v}}})] \ldots](f_H)\|_{L^2(\mathbb{R}^{\vec{d}_{\boldsymbol{v}}})} \gtrsim \|b(\vec{x}^{\,0}_{\hat{\boldsymbol{v}}})\|_{\mathrm{BMO}_{(v_1)\ldots(v_l)}}$$

where $\mathbb{R}^{\vec{d}_{\boldsymbol{v}}} = \mathbb{R}^{d_{v_1}} \times \ldots \times \mathbb{R}^{d_{v_l}}$.

We now consider the commutator with the same cones but with full symbol $b = b(\cdot, \ldots, \cdot)$. Due to the lack of action on the variables not indexed by v_s, in the commutator, we again have $[T_{v_1,D_{v_1}}, \ldots, [T_{v_l,D_{v_l}}, b] \ldots](f_Hg) = g \cdot [T_{v_1,D_{v_1}}, \ldots, [T_{v_l,D_{v_l}}, b] \ldots](f_H)$ for g that only depends upon variables not indexed by v_s. Again using that multiplication operators in L^2 have norms equal to the L^∞ norm of their symbol, for the 'worst' L^2-normalized g we have

$$
\begin{aligned}
\|[T_{v_1,D_{v_1}},\ldots,[T_{v_l,D_{v_l}},b]\ldots](f_H g)\|_{L^2(\mathbb{R}^{\vec{d}})} &= \sup_{\vec{x}_{\hat{\boldsymbol{v}}}} \|[T_{v_1,D_{v_1}},\ldots,[T_{v_l,D_{v_l}},b(\vec{x}_{\hat{\boldsymbol{v}}}^{\,0})]\ldots](f_H)\|_{L^2(\mathbb{R}^{\vec{d}_{\boldsymbol{v}}})}\\
&\geqslant \|[T_{v_1,D_{v_1}},\ldots,[T_{v_l,D_{v_l}},b(\vec{x}_{\hat{\boldsymbol{v}}}^{\,0})]\ldots](f_H)\|_{L^2(\mathbb{R}^{\vec{d}_{\boldsymbol{v}}})}\\
&\gtrsim \|b(\vec{x}_{\hat{\boldsymbol{v}}}^{\,0})\|_{\mathrm{BMO}_{(v_1)\ldots(v_l)}(\mathbb{R}^{\vec{d}_{\boldsymbol{v}}})}\\
&= \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.
\end{aligned}
$$

Note that the test function g can be chosen with well distributed Fourier transform. Take any cones in the variables not indexed by v_s and let $\vec{D}$ denote the tensor product of their projections. $f_T = P_{\vec{D}} g$. Notice that $\|[T_{v_1,D_{v_1}},\ldots,[T_{v_l,D_{v_l}},b]\ldots](f_H f_T)\|_{L^2(\mathbb{R}^{\vec{d}})} \gtrsim \|[T_{v_1,D_{v_1}},\ldots,[T_{v_l,D_{v_l}},b]\ldots](f_H g)\|_{L^2(\mathbb{R}^{\vec{d}})}$ with constants depending upon smallness of the aperture of the chosen cones. Notice that the test function $f := f_H f_T$ has the Fourier support as required in part *(2)* of the statement of Lemma 5.5.

Now build cones $\vec{D}_s$ from the D_{v_s} and the other chosen cones D_k as well as operators $T_{s,\vec{D}_s}$. Notice that the commutators $[T_{v_1,D_{v_1}},\ldots,[T_{v_l,D_{v_l}},b]\ldots]$ and $[T_{1,\vec{D}_1},\ldots,[T_{l,\vec{D}_l},b]\ldots]$ reduce significantly when applied to a test function f with Fourier support as ours. When the operators $T_{v_s,D_{v_s}}$ or any tensor product $T_{s,\vec{D}_s}$ fall directly on f, the contribution is zero due to opposing Fourier supports of the test function and the symbols of the operators. The only terms left in the commutators $[T_{1,\vec{D}_1},\ldots,[T_{l,\vec{D}_{v_l}},b]\ldots](f)$ and $[T_{v_1,D_{v_1}},\ldots,[T_{v_l,D_{v_l}},b]\ldots](f)$ have the form $\bigotimes_s T_{s,\vec{D}_s}(bf)$ and $\bigotimes_s T_{v_s,D_{v_s}}(bf)$ respectively.

By repeated use of Lemma 5.6 we have the operator norm estimates for any symbol b, valid on the subspace of functions with Fourier support as described for f: $\|\bigotimes_s T_{s,\vec{D}_s} b\|_{2\to 2} = \|\bigotimes_s T_{v_s,D_{v_s}} b\|_{2\to 2}$. We conclude the existence of a normalized test function f with Fourier support as described in the statement *(2)* of Lemma 5.5 so that $\|\bigotimes_s T_{s,\vec{D}_s}(bf)\|_2 \gtrsim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}$. We get the desired estimate

$$\|[T_{1,\vec{D}_1},\ldots,[T_{l,\vec{D}_l},b]\ldots]\|_{L^2(\mathbb{R}^{\vec{d}})\circlearrowleft} \gtrsim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})},$$

and test function f as desired. □

It seems not possible to pass directly to a lower commutator estimate in tensor products of Riesz transforms from that in tensor products of cone operators. Just using tensor products of operators adapted to cones merely gives us *some* lower bound where we are unable to control that a Riesz transform does appear in every variable such as required in *(3)* of Theorem 5.1. The reason for this will become clear as we advance in the argument. Instead of using operators $T_{s,\vec{D}_s}$ directly, we will build upon them more general multi-parameter Journé type cone operators not of tensor product type that we now describe.

Let us explain the multiplier we need for i copies of $\mathbb{S}^{d-1}$ when all dimensions are the same. We will explain how to pass to the case of i copies of varying dimension d_k below. A picture illustrating a base case can be found at the end of this section.

For $0 < b < a < 1$, let $\varphi : [-1,1] \to [-1,1]$ be a smooth function with $\varphi(x) = 1$ when $a \leqslant x \leqslant 1$ and $\varphi(x) = 0$ when $b \geqslant x \geqslant 0$. And let φ be odd, meaning antisymmetric about $t = 0$. The function φ gives rise to a zonal function with pole ξ_1 on the first copy of $\mathbb{S}^{d-1}$, denoted by $C_1(\xi_1; \eta_1)$. This is the multiplier of a one-parameter Calderón-Zygmund operator adapted to a cone $D(\xi_1, r)$ for $r = \pi/2(1-a)$. For $i > 1$ we define $C_k(\xi_1, \ldots, \xi_k; \eta_1, \ldots, \eta_k)$ for $1 < k \leqslant i$ inductively. In what follows, expectation is taken with respect to traces of surface measure. When $\eta_i = \pm\xi_i$, then conditional expectation is over a one-point set.

$$C_k(\xi_1, \ldots, \xi_k; \eta_1, \ldots, \eta_k) = \mathbb{E}_{a_{k-1}}(C_{k-1}(\xi_1, \ldots, a_{k-1}; \eta_1, \ldots, \eta_{k-1}) \mid d(a_{k-1}, \xi_{k-1}) = d(\eta_k, \xi_k)).$$

If the dimensions are not equal take $d = \max_{d_j}$ and imbed $\mathbb{S}^{d_j-1}$ into $\mathbb{S}^{d-1}$ by the map $\xi = (\xi_1, \ldots, \xi_{d_j}) \mapsto (\xi_1, \ldots, \xi_{d_j}, 0, \ldots, 0)$. Obtain in this manner the function C_i and then restrict to the original number of variables when the dimension is smaller than d.

The multiplier $C_i(\vec{\xi}; \cdot)$ gives rise to a multi-parameter Calderón-Zygmund operator of convolution type (but not of tensor product type), $\mathbf{T}_{\vec{\xi}} = \mathbf{T}_{C_i(\vec{\xi};\cdot)}$. Infact, it is defined through principal value convolution against a kernel $\mathbf{K}_{\vec{\xi}}(x_1, \ldots, x_i)$ such that

$$\forall l : \int_{\alpha < |x_l| < \beta} \mathbf{K}_{\vec{\xi}}(x_1, \ldots, x_i) dx_l = 0, \forall 0 < \alpha < \beta, x_j \in \mathbb{R}^{d_j} \text{ fixed } \forall j \neq l,$$

$$\left| \frac{\partial^{|\vec{n}|}}{\partial x_1^{n_1} \ldots \partial x_i^{n_i}} \mathbf{K}_{\vec{\xi}}(x_1, \ldots, x_i) \right| \leqslant A_{\vec{n}} |x_1|^{-d_1 - n_1} \ldots |x_i|^{-d_i - n_i}, n_j \geqslant 0.$$

This kind of operator is a special case of the more general, non-convolution type discussed in Section 4. It has many other nice features that will facilitate our passage to Riesz transforms. One of them is its very special representation in terms of homogeneous polynomials, the other one a lower estimate in terms of the $\mathrm{BMO}_{\mathcal{I}}$ norm.

Lemma 5.7. *Let C_i be a multiplier in $\bigotimes_{k=1}^{i} \mathbb{R}^{d_k}$ be as described above, with any fixed direction and aperture. Let m be an integer of order $d = \max d_k$. For any $\delta > 0$, the function C_i has an approximation by a polynomial C_i^N in the $\prod_{k=1}^{i} d_k$ variables $\{\prod_k \eta_{k,j_k} \mid 1 \leqslant k \leqslant i, 1 \leqslant j_k \leqslant d_k\}$ so that $\|C_i - C_i^N\|_{\mathcal{C}^m(\mathbb{S}^{d_k-1})} < \delta$ in each variable separately.*

$\mathcal{C}^m$ indexes the norm of uniform convergence in functions that are m times continuously differentiable. On the space side, C_i^N corresponds to an operator that is a polynomial in Riesz transforms of the variables $\bigotimes_k R_{k,j_k}$.

Lemma 5.8. *We are in $\mathbb{R}^{\vec{d}}$ with partition $\mathcal{I} = (I_s)_{1 \leqslant s \leqslant l}$. Let $\vec{\Upsilon}$ consist of vectors $\vec{\xi} = (\xi_k)_{k=1}^{t}$ with $\xi_k \in \Upsilon_k$. Let $\vec{\Upsilon}^{(s)}$ consist of $\vec{\xi}^{(s)} = (\xi_k)_{k \in I_s}$. Let us consider the class of Journé type cone multipliers $\boldsymbol{D}_s = C_{i_s}(\vec{\xi}^{(s)}; \cdot)$ of aperture r_s with associated multi-parameter Calderón-Zygmund* *operators* $\boldsymbol{T}_{s,\boldsymbol{D}_s}$. *Then we have the two-sided estimate*

$$\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{\xi}\in\vec{\Upsilon}} \|[\,\boldsymbol{T}_{1,\boldsymbol{D}_1},\ldots,[\,\boldsymbol{T}_{l,\boldsymbol{D}_l},b]\ldots]\|_{L^2(\mathbb{R}^{\vec{d}})\circlearrowleft} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

In order to proceed with the proof of these lemmas, we will need some well known facts about spherical harmonics and zonal harmonics in particular, that we now review briefly. Fix a pole $\xi\in\mathbb{S}^{d-1}$. The zonal harmonic with pole ξ of degree n is written as $Z^{(n)}_{\xi}(\eta)$. With $t=\langle\xi,\eta\rangle\in[-1,1]$, one writes $Z^{(n)}_{\xi}(\eta)=P_n(t)$ where P_n is the Legendre polynomial of degree n. It is common to suppress the dependence on d in the notation for $Z^{(n)}_{\xi}$ and P_n. The Rodrigues formula with a specific weight useful for a change of variables on the sphere $\mathbb{S}^{d-1}$ is

$$P_n(t)=\frac{(-1)^n}{2^n(n+(d-3)/2)_n}(1-t^2)^{(3-d)/2}\left(\frac{\partial}{\partial t}\right)^n(1-t^2)^{n+(d-3)/2}.$$

Here $(k)_n$ denotes the descending factorial $k(k-1)\ldots(k-n+1)$. Writing Ω_{d-1} for the measure of $\mathbb{S}^{d-1}$, we have orthogonality by the formulas

$$0=\int_{\mathbb{S}^{d-1}} Z^{(n)}_{\xi}(\eta)Z^{(m)}_{\xi}(\eta)d\sigma_{d-1}=\Omega_{d-1}\int_{-1}^{1}P_n(t)P_m(t)(1-t^2)^{(d-3)/2}dt$$

and normalization

$$\int_{-1}^{1}P_n^2(t)(1-t^2)^{(d-3)/2}dt=\frac{\Omega_{d-1}}{N(d,n)\Omega_{d-2}}$$

where $N(d,n)$ is the dimension of the space of homogeneous harmonic polynomials of degree n in d variables. We have $P_n(1)=1$, $P_n(-1)=(-1)^nP_n(1)$ and $|P_n(t)|\leqslant 1$ for $t\in[-1,1]$.

$Z^{(n)}_{\xi}$ are reproducing and for any spherical harmonic of degree n, $Y^{(n)}$, we have

$$Y^{(n)}(\xi)=\frac{N(d,n)}{\Omega_{d-1}}\int_{\mathbb{S}^{d-1}}Y^{(n)}(\eta)Z^{(n)}_{\xi}(\eta)d\sigma_{d-1} \tag{4}$$

as well as for any function φ zonal with respect to the pole ξ

$$\int_{\mathbb{S}^{d-1}}\varphi(\langle\xi,\eta\rangle)Y^{(n)}(\eta)d\sigma_{p-1}=\Omega_{d-2}Y^{(n)}(\xi)\int_{-1}^{1}\varphi(t)P_n(t)(1-t^2)^{(d-3)/2}dt. \tag{5}$$

The addition theorem for Legendre polynomials says that for any complete system $\{Y^{(n)}_k\}$ of spherical harmonics of degree n:

$$Z^{(n)}_{\xi}(\eta)=\frac{\Omega_{d-1}}{N(d,n)}\sum_{k=1}^{N(d,n)}Y^{(n)}_k(\xi)Y^{(n)}_k(\eta). \tag{6}$$

When $Y^{(n)}$ is harmonic and homogeneous of degree n with $Y^{(n)}(\xi)=1$ and $Y^{(n)}(R\eta)=Y^{(n)}(\eta)$ for any rotation $R\in\mathcal{O}(d)$ with $R\xi=\xi$, then $Y^{(n)}=Z^{(n)}_{\xi}$, the zonal harmonic of degree n with pole ξ.

The lemma below will aid us in understanding the special form of the functions C_i.

Lemma 5.9. *Let $\xi_1, \xi_2 \in \mathbb{S}^{d-1}$. We have*

$$\mathbb{E}_{a_1}(Z^{(n)}_{\eta_1}(a_1) \mid d(\xi_1, a_1) = d(\xi_2, \eta_2)) = \mathbb{E}_{a_2}(Z^{(n)}_{\eta_2}(a_2) \mid d(\xi_2, a_2) = d(\xi_1, \eta_1)) = Z^{(n)}_{\xi_1}(\eta_1) Z^{(n)}_{\xi_2}(\eta_2).$$

Proof. The first equality is a change of variable, thanks to symmetry of the zonal harmonic in its variables and invariance with respect to action of the measure preserving elements of the orthogonal group fixing poles ξ_1 or ξ_2, that we now detail. By a rotation in one of the spheres, assume $\xi_1 = \xi_2 = \xi$. Take a small ball

$$B_{\xi,\eta_1}(a_2^0; \varepsilon_2) = \{a_2 : d(a_2, a_2^0) < \varepsilon_2\} \cap \{a_2 : d(a_2, \xi) = d(\eta_1, \xi)\}.$$

Note $\{a_2 : d(a_2, \xi) = d(\eta_1, \xi)\} \sim \mathbb{S}^{d-2}$. Every $a_2 \in B_{\xi,\eta_1}(a_2^0; \varepsilon_2)$ gives rise to a canonical orthogonal map σ_{a_2} along geodesics in a scaled copy of $\mathbb{S}^{d-2}$. Lifted to $\mathbb{S}^{d-1}$, these are orthogonal maps fixing ξ. Let σ^0 fix ξ and map a_2^0 to η_1. Let $a_1^0 = \sigma^0(\eta_2)$. We observe that $\{\sigma^0 \sigma_{a_2}(\eta_2) : a_2 \in B_{\xi,\eta_1}(a_2^0; \varepsilon_2)\} = B_{\xi,\eta_2}(a_1^0; \varepsilon_1)$ with ε_1 so that

$$\mathbb{P}(d(a_2, a_2^0) < \varepsilon_2 \mid d(\xi, a_2) = d(\xi, \eta_1)) = \mathbb{P}(d(a_1, a_1^0) < \varepsilon_1 \mid d(\xi, a_1) = d(\xi, \eta_2)).$$

Together with the symmetry and the rotation property $Z^{(n)}_\eta(a) = Z^{(n)}_a(\eta) = Z^{(n)}_{\sigma(a)}(\sigma(\eta))$, we obtain the first equality.

For fixed a_1, the function $Z^{(n)}_{\eta_1}(a_1) = Z^{(n)}_{a_1}(\eta_1)$ is a function harmonic in $\mathbb{R}^d$, n-homogeneous. These properties are preserved when taking expectation in a_1. So the expression $\mathbb{E}(Z^{(n)}_{\eta_1}(a_1) \mid d(\xi_1, a_1) = d(\xi_2, \eta_2))$ remains harmonic (regarded as a function in $\mathbb{R}^d$), n-homogeneous. From the form $\mathbb{E}(Z^{(n)}_{\eta_2}(a_2) \mid d(\xi_2, a_2) = d(\xi_1, \eta_1))$ we learn that its restriction to $\mathbb{S}^{d-1}$ depends only upon $d(\xi_1, \eta_1)$. This implies that it is a constant multiple of the zonal harmonic with pole ξ_1. Exchanging the roles of η_1 and η_2 gives

$$\mathbb{E}(Z^{(n)}_{\eta_1}(a_1) \mid d(\xi_1, a_1) = d(\xi_2, \eta_2)) = c_n Z^{(n)}_{\xi_1}(\eta_1) Z^{(n)}_{\xi_2}(\eta_2).$$

When assuming the normalization $Z^{(n)}_\xi(\xi) = 1$ then $c_n = 1$. This is a gernalisation of the classical symmetrising of the cosinus sum formula $1/2(\cos(x+y) - \cos(x-y)) = \cos(x)\cos(y)$. □

Let us now proceed with the proof of one of our lemmas

Proof. (of Lemma 5.7) It is well known that zonal harmonic series have convergence properties when representing smooth zonal functions similar to that of the Fourier transform. For any given m and sufficiently smooth φ of the type described above, then

$$C_1(\xi_1; \eta_1) = \sum_n \varphi_n Z^{(n)}_{\xi_1}(\eta_1)$$

where the convergence is $\mathcal{C}^m$-uniform. The degree of smoothness required for φ to obtain convergence in the $\mathcal{C}^m$ in the above expression depends upon m and the dimension d. For our purpose, we choose $m \geqslant d$. The coefficients φ_n are explicit by formula 5.

Let us denote this function's representation of degree N by a series of zonal harmonics by $C_1^{(N)}(\xi_1;\eta_1)$.

$$C_1^{(N)}(\xi_1;\eta_1) = \sum_{n\leqslant N} \varphi_n Z_{\xi_1}^{(n)}(\eta_1).$$

For every $\delta > 0$ there exists N so that we have the estimate

$$\|C_1^{(N)}(\xi_1;\eta_1) - C_1(\xi_1;\eta_1)\|_{\mathcal{C}^m(\mathbb{S}^{d_1-1})} < \delta.$$

In the case of i copies of spheres, we define inductively in the same manner as for C_i. Let us for the moment make all dimensions equal by the imbedding discussed above.

$$C_k^{(N)}(\xi_1,\ldots,\xi_k;\eta_1,\ldots,\eta_k) = \mathbb{E}_{a_{k-1}}(C_{k-1}^{(N)}(\xi_1,\ldots,a_{k-1};\eta_1,\ldots,\eta_{k-1}) \mid d(a_{k-1},\xi_{k-1}) = d(\eta_k,\xi_k))$$

We claim the identity

$$C_i^{(N)}(\vec{\xi};\eta_1,\eta_2,\ldots,\eta_i) = \sum_{n\leqslant N} \varphi_n \prod_{k=1}^{i} Z_{\xi_k}^{(n)}(\eta_k). \tag{7}$$

This is trivially true for $i = 1$. For $i > 1$ induct on the number of parameters using Lemma 5.9:

$$\begin{aligned}
&C_i^{(N)}(\vec{\xi};\eta_1,\ldots,\eta_i)\\
= \;&\mathbb{E}_{a_{i-1}}(C_{i-1}(\xi_1,\xi_2,\ldots,a_{i-1};\eta_1,\ldots,\eta_{i-1}) \mid d(a_{i-1},\xi_{i-1}) = d(\eta_i,\xi_i))\\
= \;&\mathbb{E}_{a_{i-1}}\left(\sum_{n\leqslant N} \varphi_n \prod_{k=1}^{i-1} Z_{\xi_k}^{(n)}(\eta_k) \mid d(a_{i-1},\xi_{i-1}) = d(\eta_i,\xi_i)\right)\\
= \;&\sum_{n\leqslant N} \varphi_n \prod_{k=1}^{i-2} Z_{\xi_k}^{(n)}(\eta_k)\mathbb{E}_{a_{i-1}}(Z_{\xi_{i-1}}^{(n)} \mid d(a_{i-1},\xi_{i-1}) = d(\eta_i,\xi_i))\\
= \;&\sum_{n\leqslant N} \varphi_n \prod_{k=1}^{i} Z_{\xi_k}^{(n)}(\eta_k).
\end{aligned}$$

The first equality is the definition of $C_i^{(N)}$, the second one is the induction hypothesis and the last an application of Lemma 5.9.

It follows that neither C_i nor $C_i^{(N)}$ depend on the order chosen in their definition and

$$C_i(\vec{\xi};\eta_1,\ldots,\eta_i) = \sum_{n} \varphi_n \prod_{k=1}^{i} Z_{\xi_k}^{(n)}(\eta_k)$$

where the convergence is in $\mathcal{C}^m$ in each variable.

Next, we study the terms arising in multipliers of the form $C_i^{(N)}$. When all dimensions are equal, indeed, $\prod_{k=1}^{i} Z_{\xi_k}^{(n)}(\eta_k)$ has the important property that, as a product of n-homogeneous polynomials, has only terms of the form

$$\prod_{k=1}^{i} \eta_k^{\alpha_k} = \prod_{k=1}^{i} \left(\prod_{j_k=1}^{d} \eta_{k,j_k}^{\alpha_{k,j_k}} \right) \tag{8}$$

where $\eta_k \in \mathbb{S}^{d-1}$ and $\alpha_k = (\alpha_{k,j_k})$ are multi-indices with $|\alpha_k| = \sum_{j_k} \alpha_{k,j_k} = n$ for all k. This form is inhertited by $C_i^{(N)}$ with varying n. It shows that $C_i^{(N)}$ is indeed a polynomial in the variables $\prod_{k=1}^{i} \eta_{k,j_k}$. When the dimensions are not equal, observe that by restricting to actually arising number of variables we certainly lose harmonicity but not n-homogeneity or the required form of our polynomials. $\square$

Proof. (of Lemma 5.8) By Lemma 5.5 we know that for each parameter $1 \leqslant s \leqslant l$ there exists a tensor product of cones $\vec{D}_s = \bigotimes_{k\in I_s} D(\xi_k, r_k)$ with $r_s := \sum_{k\in I_s} r_k < \pi/2$ and $\xi_k \in \Upsilon_k$ and test functions f_s supported as described in Lemma 5.5 part *(2)* so that

$$\|[T_{1,\vec{D}_1}, \dots, [T_{l,\vec{D}_l}, b] \dots](f)\|_2 \gtrsim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}$$

where $f = \bigotimes_{s=1}^{l} f_s$. We make a remark about the apertures r_s. Let $d(\cdot,\cdot)$ denote geodesic distance on $\mathbb{S}^{d-1}$, where antipodal points have distance π. Let $\vec{\xi}^{(s)}$ be the set of directions of $\vec{D}_s$. Remember that according to Lemma 5.5, one component had a specific direction $\xi_v^{(s)} \in \Upsilon_v$ and possibly large aperture with $(1+\tau) r_v^{(s)} < \pi/2$. Let us choose the other directions arbitrarily but with apertures $r_{\hat{v}}^{(s)}$ small enough so that $(1+\tau)(r_v^{(s)} + (i-1) r_{\hat{v}}^{(s)}) < \pi/2$. Now choose an aperture $r_s < \pi/2$ so that $(1+\tau)(r_v^{(s)} + (i-1) r_{\hat{v}}^{(s)}) < r_s < \pi/2$.

Writing $i_s = |I_s|$, we find Journé type cone multipliers $\mathbf{D}_s = C_{i_s}(\vec{\xi}^{(s)}; \cdot)$ according to the construction above with center $\vec{\xi}^{(s)}$ and aperture r_s.

We are going to observe now that $\mathbf{D}_s \equiv 1$ on $\mathrm{spt}(\vec{D}_s)$ and $\mathbf{D}_s \equiv -1$ on the Fourier support of f_s. Let us drop the dependence on s for the moment. We see in an inductive manner that $C_i(\vec{\xi}; \cdot)$ takes the value 1 in a certain ℓ^1 ball of radius $r < \pi/2$ centered at $\vec{\xi}$. We show that

$$\sum_k d(\xi_k, \eta_k) < r \Rightarrow C_i(\vec{\xi}, \eta_1, \dots, \eta_i) = 1.$$

When $i = 1$, the assertion is obviously true: $d(\xi_1, \eta_1) < r \Rightarrow C_1(\xi_1; \eta_1) = 1$ by the choice of φ, r and definition of C_1. For $i > 1$, we proceed by induction. Assume that $\sum_{k=1}^{i-1} d(\xi_k, n_k) < r \Rightarrow C_{i-1}(\xi_1, \dots, \xi_{i-1}; \eta_1, \dots, \eta_{i-1}) = 1$. Let us assume that $\sum_{k=1}^{i} d(\xi_k, \eta_k) < r$. Remembering the definition of $C_i(\vec{\xi}; \cdot)$ we assume $d(a_{i-1}, \xi_{i-1}) = d(\eta_i, \xi_i)$. By the triangle inequality $\sum_{k=1}^{i-2} d(\xi_k, \eta_k) + d(a_{i-1}, \eta_{i-1}) \leqslant \sum_{k=1}^{i-2} d(\xi_k, \eta_k) + d(a_{i-1}, \xi_{i-1}) + d(\xi_{i-1}, \eta_{i-1}) = \sum_{k=1}^{i} d(\xi_k, \eta_k) < r$. So

$$C_{i-1}(\xi_1, \xi_2, \dots, a_{i-1}; \eta_1, \dots, \eta_{i-1}) = 1$$

for all a_{i-1} we condition by. The statement for i follows.

Since $C_i(\vec{\xi};\cdot)$ does not depend upon the order of the variables in its construction, we are also able to see exactly as done above that when $\sigma_k = -1$ for exactly one choice of k, then $\sum_k d(\sigma_k \xi_k, \eta_k) < r \Rightarrow C_i(\vec{\xi};\eta_1,\ldots,\eta_i) = -1$.

Consider associated multi-parameter Calderón-Zygmund operators $\mathbf{T}_{s,\mathbf{D}_s}$ and $\vec{\mathrm{Id}}_s = \bigotimes_{k\in I_s} \mathrm{Id}_k$ and Id_k the identity on the k^{th} variable. Now

$$\begin{aligned}[\mathbf{T}_{1,\mathbf{D}_1},\ldots,[\mathbf{T}_{l,\mathbf{D}_l},b]\ldots](f) &= [\mathbf{T}_{1,\mathbf{D}_1}+\vec{\mathrm{Id}}_1,\ldots,[\mathbf{T}_{l,\mathbf{D}_l}+\vec{\mathrm{Id}}_l,b]\ldots](f)\\ &= \bigotimes_{s=1}^{l}(\mathbf{T}_{s,\mathbf{D}_s}+\vec{\mathrm{Id}}_s)(bf)\end{aligned}$$

With $\|\bigotimes_{s=1}^{l}(\mathbf{T}_{s,\mathbf{D}_s}+\vec{\mathrm{Id}}_s)(bf)\|_2 \geqslant \|\bigotimes_{s=1}^{l} T_{s,\vec{D}_s}(bf)\|_2 = \|[T_{1,\vec{D}_1},\ldots,[T_{l,\vec{D}_l},b]\ldots](f)\|_2 \gtrsim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}$ we get the desired lower bound

$$\|[\mathbf{T}_{1,\mathbf{D}_1},\ldots,[\mathbf{T}_{l,\mathbf{D}_l},b]\ldots]\|_{L^2(\mathbb{R}^{\vec{d}})\circlearrowleft} \gtrsim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

□

Let us illustrate the base case $(\mathbb{S}^1)^2$ by a picture. The picture is simplified in the sense that the odd function φ above is replaced by an indicator function of an interval.

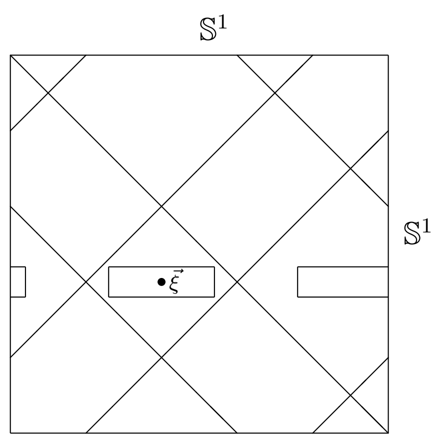

Figure 1: $\mathbb{R}^2 \times \mathbb{R}^2$

Cone functions based on the oblique strips containing $\vec{\xi}$ are averaged. In the two-dimensional case, $\mathbb{S}^1$, expectation is over a one or two point set only. The rectangle around $\vec{\xi}$ with sides parallel to the axes illustrates the support of the tensor product of cone operators with direction $\vec{\xi}$. The longer side is the aperture that arises from the Hankel part [17]. The short sides can be chosen freely as they arise from the Toeplitz part and is chosen small so that the rectangle fits into the oblique square. The other small rectangle corresponds to the Fourier support of the test function f.

Now, we are ready to prove our main theorem.

Proof. (of Theorem 5.1)

Different from the Hilbert transform case, both lower bounds require separate proofs. This is a notable difference that stems from the loss of orthogonal subspaces in conjunction with the special form of the Hilbert transform only seen in one variable. It does not seem possible to get a lower estimate *(3)⇒(2)* directly.

(1)⇔(2) The upper bound *(1)⇒(2)* is trivial and follows like the estimate 4.1. The lower bound *(2)⇒(1)* follows from Wiener's lemma in combination with the main result in [17],

the two-sided estimate for iterated commutators with Riesz transforms, similar to the first arguments used in 5.5.

(1)⇔(3) The upper bound *(1)⇒(3)* follows from 4.1. The lower bound *(3)⇒(1)* uses the considerations leading up to this proof: Suppressing again the dependence on s, we see that the multiplier C_i is an odd, smooth, bounded function in each η_k when the other variables are fixed. Furthermore, since φ, written as a function of $t = \langle \xi, \eta \rangle$ is odd with respect to $t = 0$, then the above series has $\varphi_n \neq 0$ at most when n is odd and thus $Z_\xi^{(n)}$ is odd. So $C_i^{(N)}$ is as a sum of odd functions odd.

We are now also ready to see that $\mathbf{T}_{\vec{\xi}}$, the Journé operator associated to the cone $C_i(\vec{\xi}; \cdot)$ as well as the operator associated to $C_i^{(N)}(\vec{\xi}; \cdot)$ are paraproduct free. Infact, applied to a test function $f = \bigotimes_k f_k$ with f_k acting on the k^{th} variable and where $f_l \equiv 1$ for some l gives $\mathbf{T}_{\vec{\xi}}(f) = 0$. To see this, apply the multiplier $C_i^{(N)}(\vec{\xi}; \cdot)$ in the l variable (acting on 1) first, leaving the other Fourier variables fixed. The multiplier function

$$\eta_l \mapsto C_i^{(N)}(\vec{\xi}; \eta_1, \ldots, \eta_i) = \sum_{n \leqslant N} \varphi_n Z_{\xi_l}^{(n)}(\eta_l) \prod_{k \neq l, k=1}^{i} Z_{\xi_k}^{(n)}(\eta_k)$$

is, as a sum of odd functions, odd on $\mathbb{S}^{d_l - 1}$, bounded by 1 and uniformly smooth for all choices of η_k with $k \neq l$. As such it gives rise to a paraproduct-free convolution type Calderón-Zygmund operator in the l variable whose values are multi-parameter multiplier operators.

Due to the convergence properties proved above, the difference

$$C_i(\vec{\xi}; \cdot) - C_i^{(N)}(\vec{\xi}; \cdot)$$

gives rise to a paraproduct free Journé operator with Calderón-Zygmund norm depending on N. This is seen by an application of an appropriate version of the Marcinkievicz multiplier theorem.

By our stability result, Corollary 4.5, there exist integers N_s so that $C_s^{(N_s)}(\vec{\xi}_s; \cdot)$ with $\xi_k \in \Upsilon_k$ is a characterising set of operators via commutators for $\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$. This is a finite set of possibilities because of the universal choice of the r_s and finiteness of the set $\vec{\Upsilon}$. Using the multi-parameter analog of the observation $[AB, b] = A[B, b] + [A, b]B$ and the special form of the $C_s^{(N_s)}(\vec{\xi}; \cdot)$, leaves us with the desired lower bound: Observe that when $[AB, b]$ has large L^2 norm then either $[A, b]$ or $[B, b]$ has a fair share of the norm. We use this argument finitely many times in a row for operators that are polynomials in tensor products in Riesz transforms $\bigotimes_{k \in I_s} R_{k, j_k}$. This finishes (*(3)⇒(1)*). □

6. Weak Factorization

It is well known, that theorems of this form have an equivalent formulation in the language of weak factorization of Hardy spaces. We treat the model case $\mathbb{R}^{\vec{d}} = \mathbb{R}^{(d_1, d_2, d_3)}$ and

$\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})$ only for sake of simplicity. The other statements are an obvious generalization. For the corresponding collections of Riesz transforms $\mathcal{R}_{k,j_k}$ and $b \in \mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})$, $1 \leqslant s \leqslant 3$, by unwinding the commutator one can define the operator $\Pi_{\vec{j}}$ such that

$$\langle [R_{2,j_2}, [R_{1,j_1} R_{3,j_3}, b]] f, g \rangle_{L^2} = \langle b, \Pi_{\vec{j}}(f, g) \rangle_{L^2}.$$

Consider the Banach space $L^2 * L^2$ of all functions in $L^1(\mathbb{R}^{\vec{d}})$ of the form $f = \sum_{\vec{j}} \sum_i \Pi_{\vec{j}}(\phi_i^{\vec{j}}, \psi_i^{\vec{j}})$ normed by

$$\|f\|_{L^2 * L^2} = \inf\{\sum_{\vec{j}} \sum_i \|\phi_i^{\vec{j}}\|_2 \|\psi_i^{\vec{j}}\|_2\}$$

with the infimum running over all possible decompositions of f. Applying a duality argument and the two-sided estimate in Corollary 5.2 we are going to prove the following weak factorization theorem.

Theorem 6.1. *$H^1_{Re}(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1_{Re}(\mathbb{R}^{(d_2,d_3)})$ coincides with the space $L^2 * L^2$. In other words, for any $f \in H^1_{Re}(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1_{Re}(\mathbb{R}^{(d_2,d_3)})$ there exist sequences $\phi_i^{\vec{j}}, \psi_i^{\vec{j}} \in L^2$ such that $f = \sum_{\vec{j}} \sum_i \Pi_{\vec{j}}(\phi_i^{\vec{j}}, \psi_i^{\vec{j}})$ and $\|f\| \sim \sum_{\vec{j}} \sum_i \|\phi_i^{\vec{j}}\|_2 \|\psi_i^{\vec{j}}\|_2$.*

Proof. Let's first show that $L^2 * L^2$ is a subspace of $H^1_{\mathrm{Re}}(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1_{\mathrm{Re}}(\mathbb{R}^{(d_2,d_3)})$. Recalling the remark after Theorem 2.3, this is the same as to show $\forall f \in L^2 * L^2$, f is a bounded linear functional on $\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})$. This follows from the upper bound on the commutators since

$$\langle b, \sum_{\vec{j}} \sum_i \Pi_{\vec{j}}(\phi_i^{\vec{j}}, \psi_i^{\vec{j}}) \rangle = \sum_{\vec{j}} \sum_i \langle [R_{2,j_2}, [R_{1,j_1} R_{3,j_3}, b]] \phi_i^{\vec{j}}, \psi_i^{\vec{j}} \rangle.$$

Now we are going to show

$$\sup_{f \in L^2 * L^2} \left\{ \left| \int f b \right| : \|f\|_{L^2 * L^2} \leqslant 1 \right\} \approx \|b\|_{\mathrm{BMO}_{(13)(2)}}$$

which gives the equivalence of $H^1_{\mathrm{Re}}(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1_{\mathrm{Re}}(\mathbb{R}^{(d_2,d_3)})$ norm and the $L^2 * L^2$ norm, thus showing that the two spaces are the same.

To see this, note that the direction $\lesssim$ is trivial, and the direction $\gtrsim$ is implied by the lower bound of commutators. For any $b \in \mathrm{BMO}_{(13)(2)}$, there exists $\vec{j}$ such that $\|b\|_{\mathrm{BMO}_{(13)(2)}} \lesssim \|[R_{2,j_2}, [R_{1,j_1} R_{3,j_3}, b]]\|$. Hence, there exist $\phi, \psi \in L^2$ with norm 1 such that

$$\|b\|_{\mathrm{BMO}_{(13)(2)}} \lesssim |\langle [R_{2,j_2}, [R_{1,j_1} R_{3,j_3}, b]] \phi, \psi \rangle| = |\langle b, \Pi_{\vec{j}}(\phi, \psi) \rangle| \leqslant LHS,$$

which completes the proof. □

7. Remarks about our results in L^p

As mentioned before, the two-sided estimates stated in section 5 and in particular Theorem 5.3 hold for all $1 < p < \infty$. The fact that upper estimates hold in L^p in the case where no tensor products are present is proved in [17] as well as [18]. It stems from the fact that endpoint estimates for multi-parameter paraproducts hold for all $1 < p < \infty$ [22], [23]. The tensor product case can be derived following the argument for Theorem 4.1. The lower estimate or the necessity of the BMO condition can be derived from interpolation. Infact, suppose we have uniform boundedness of our commutators with operators running through all choices of Riesz transforms and some symbol b in L^p. Then by duality, we have boundedness in L^q where $1/p + 1/q = 1$. In fact, $[T,b]^*f = -[T^*,\bar{b}]f = -\overline{[\overline{T^*},b]\bar{f}}$ shows that the boundedness of adjoints is inherited. The same reasoning holds for iterated commutators of tensor products. Thus by interpolation, the boundedness holds in L^2 and the symbol function b necessarily belongs to the required BMO class.

References

[1] Arlen Brown and Paul Halmos, *Algebraic properties of Toeplitz operators*, J. Reine Angew. Math. **213** (1964), 89–102. ↑9

[2] Lennart Carleson, *A counterexample for measures bounded on H^p spaces for the bidisk*, Mittag-Leffler Rep. No. 7, Inst. Mittag-Leffler (1974). ↑2, 5

[3] Sun-Yung Chang and Robert Fefferman, *Some recent developments in Fourier analysis and H^p-theory on product domains*, Bull. Amer. Math. Soc. (N.S.) **12** (1985), no. 1, 1–43.MR 86g:42038 ↑2, 4

[4] ______, *A continuous version of duality of H^1 with BMO on the bidisc*, Ann. of Math. (2) **112** (1980), no. 1, 179–201. ↑2, 4

[5] Ronald Coifman, Richard Rochberg, and Guido Weiss, *Factorization theorems for Hardy spaces in several variables*, Ann. of Math. (2) **103** (1976), no. 3, 611–635. ↑2, 3, 11

[6] Misha Cotlar and Cora Sadosky, *The Helson-Szegö theorem in L^p of the bidimensional torus*, Contemp. Math. **107** (1990), 19–37. ↑5

[7] Laurent Dalenc and Yumeng Ou, *Upper bound for multi-parameter iterated commutators*, Preprint (2014). ↑2, 11, 12, 14, 15, 16, 19, 21

[8] Laurent Dalenc and Stefanie Petermichl, *A lower bound criterion for iterated commutators*, J. Funct. Anal. **266** (2014), no. 8, 5300–5320. ↑23, 24

[9] Robert Fefferman, *Bounded mean oscillation on the polydisk*, Ann. of Math. (2) **110** (1979), no. 2, 395–406.MR 81c:32016 ↑5

[10] ______, *Singular integrals on product domains*, Bull. Amer. Math. Soc. (N.S.) **4** (1981), no. 2, 195–201. ↑12

[11] ______, *Harmonic analysis on product spaces*, Ann. of Math. (2) **126** (1987), no. 1, 109–130.MR 90e:42030 ↑5

[12] Sarah Ferguson and Michael Lacey, *A characterisation of product BMO by commutators*, Acta Math. **189** (2002), no. 2, 143–160. ↑2

[13] Sarah Ferguson and Cora Sadosky, *Characterisations of bounded mean oscillation on the polydisk in terms of Hankel operators and Carleson measures*, J. Anal. Math. **81** (2000), 239–267. ↑1, 5

[14] Tuomas Hytönen, *The sharp weighted bound for general Calderón-Zygmund operators*, Ann. of Math. **175** (2012), no. 3, 1473–1506. ↑11

[15] Jean-Lin Journé, *Calderón-Zygmund operators on product spaces*, Rev. Mat. Iberoamericana **1** (1985), no. 3, 55–91. ↑4, 12

[16] ______, *A covering lemma for product spaces*, Proc. Amer. Math. Soc. **96** (1986), no. 4, 593–598. ↑2

[17] Michael Lacey, Stefanie Petermichl, Jill Pipher, and Brett Wick, *Multiparameter Riesz commutators*, Amer. J. Math. **131** (2009), no. 3, 731–769. ↑2, 3, 5, 11, 23, 32, 35

[18] ______, *Iterated Riesz commutators: a simple proof of boundedness.*, Contemp. Math. **505** (2010), 171–178. ↑11, 35

[19] Michael Lacey and Erin Terwilleger, *Hankel operators in several complex variables and product BMO*, Houston J. Math. **35** (2009), no. 1, 159–183. ↑2

[20] Henri Martikainen, *Representation of bi-parameter singular integrals by dyadic operators*, Adv. Math. **229** (2012), no. 3, 1734–1761. ↑4, 12, 13

[21] Henri Martikainen and Tuomas Orponen, *Some obstacles in characterising the boundedness of bi-parameter singular integrals*, arXiv:1404.2216, 1–10. ↑12

[22] Camil Muscalu, Jill Pipher, Terence Tao, and Christoph Thiele, *Bi-parameter paraproducts*, Acta Math. **193** (2004), no. 2, 269–296. ↑35

[23] ______, *Multi-parameter paraproducts*, Rev. Mat. Iberoam. **22** (2006), no. 3, 963–976. ↑35

[24] Zeev Nehari, *On bounded bilinear forms*, Ann. of Math. (2) **65** (1957), 153–162. ↑1

[25] Yumeng Ou, *A $T(b)$ theorem on product spaces*, arXiv:1305.1691, to appear Trams. Amer. Math. Soc., 1–45. ↑12

[26] ______, *Multi-parameter singular integral operators and representation theorem*, arXiv:1410.8055, 1–28. ↑4, 12, 21

[27] Stefanie Petermichl, *Dyadic Shifts and a Logarithmic Estimate for Hankel Operators with Matrix Symbol*, Comptes Rendus Acad. Sci. Paris **1** (2000), no. 1, 455–460. ↑11

[28] Stefanie Petermichl, Sergei Treil, and Alexander Volberg, *Why the Riesz transforms are averages of the dyadic shifts?*, Publicacions matematiques, ISSN 0214-1493 **46** (2002), no. 2, 209–228. ↑11

[29] Jill Pipher, *Journé's covering lemma and its extension to higher dimensions*, Duke Math. J. **53** (1986), no. 3, 683–690. ↑2

[30] Sandra Pott and Paco Villarroya, *A T(1) theorem on product spaces.* ↑12

Yumeng Ou, Brown University, 151 Thayer Street, Providence RI 02912, USA
E-mail address: yumeng_ou@brown.edu

Stefanie Petermichl, Université Paul Sabatier, Institut de Mathématiques de Toulouse, 118 route de Narbonne, F-31062 Toulouse, France
E-mail address: stefanie.petermichl@gmail.com

Elizabeth Strouse, Université de Bordeaux, Institut de Mathématiques de Bordeaux UMR 5251, 351 cours de la Libération, F-33405 Talence, France
E-mail address: elizabeth.strouse@math.u-bordeaux1.fr